Susceptibility To Late Onset Hearing Loss: An Investigation Into Genetic Variation At The Brn-3c Locus.

Lisa Sarah Nolan

A thesis submitted for the Degree of Doctor of Philosophy

Department of Immunology and Molecular Pathology

University College London

2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

'I, Lisa Sarah Nolan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

29th Sept 2006.

Abstract

Brn-3c (*Brn3.1, POU4F3*) encoding a POU domain transcription factor is a candidate gene for late onset sensorineural hearing loss, which is exhibited by a large proportion of the ageing population. To identify common sequence variants at the *Brn-3c* locus mutation scanning of the *Brn-3c* cDNA, intron and 5'-flanking region was performed by PCR-SSCP analysis in 45 members of the general population. Seven polymorphic sites were identified of which five within the *Brn-3c* 5'-flanking region appear common. A functional screening approach utilising *in-vitro* assays suggests that at least three common sequence variants in the *Brn-3c* 5'-flanking region could have a functional affect: -566$(GT)_{17\text{-}23}$, -1391A>C and a complex multi-allelic poly-G polymorphism at -3432 that exhibits multiple variations in length together with single base substitutions within the guanine repeat. The -3432poly-G polymorphism modifies the binding affinity of an OC-2 derived nuclear protein and there is convincing evidence that this is the transcription factor SP1. Use of purified human recombinant SP1 protein, *in-vitro* translated SP1 and *in-vitro* translated SP3 confirms that the -3432polyG polymorphism modulates a high affinity SP family binding site and evidence suggests that this alters the regulation of the *Brn-3c* promoter when SP1 levels are limiting, $p<0.05$. Moreover, the data suggest a functional interaction between the -3432poly-G polymorphism and the -566$(GT)_{17\text{-}23}$ repeat which associate to determine the response of the *Brn-3c* gene to SP1. Similarly, evidence suggests that the variant allele, -1391C has a reduced affinity for an OC-2 derived nuclear protein and this is consistent with a significant decrease in basal activity of the *Brn-3c* promoter, $p<0.05$. Both -3432poly-G polymorphism and -1391A>C were genotyped for a pilot association study but allelic frequencies were not found to significantly differ between the patient and control populations examined (by χ^2 analysis). Further large-scale population studies are required to establish whether these common sequence variants are associated with late onset hearing loss.

Dedication

I would like to dedicate this thesis to a much loved nan, Vera Elford who told me:

"Something good is always round the corner"

1929 – 1999

Acknowledgments

Firstly, I would like to thank my supervisor Dr Sally Dawson for giving me the opportunity to undertake this PhD project and for all her words of advice and patience during this project and preparation of this thesis. I would also like to extend special thanks to her for adopting the role of my 'note-taker' at many meetings and seminars. Secondly, I would like to thank Research Into Ageing, The Teresa Rosenbaum Golden Charitable Trust and Deafness Research UK for funding this PhD studentship. I would also like to give special thanks to Dr Barbara Cadge for collaborating on this project and for all her time spent setting up sample recruitment from the adult hearing aid clinic at the Royal Free Nose, Ear and Throat Hospital, London, U.K. I would also like to extend special thanks to my colleagues within the Auditory Research Group at the Windeyer Institute: Richa Sud and Sanjay Jagutpal for their friendship, their words of advice and for making the many hours spent working in the lab an enjoyable place to be! I would especially like to thank Richa for supplying copious amounts of chocolate on 'Fridays' (in the office not in the lab!). Importantly, I could not write this acknowledgment without mentioning some very dear and old (sorry – well we are 30 now!) friends of mine: Sara Hall, Cee Matthews and Nicola Harvey for all their support over the last few years. Sara for keeping me sane and being there to listen to me talk and yet talk some more about my work, my thesis, my life (!) over many bottles of red wine (– how I love Fridays in London!); Cee and Nicola for keeping me smiling with their words of encouragement and being such great friends. *Most importantly*, I could not have achieved this without the love and solid support of my family: my parents Michael and Joyce Nolan, my grandfather Rowland Elford and Darryl Enfield, my partner, who has been an absolute rock of support to me; I would like to thank him for help giving me the strength to believe in myself.

L. Nolan, London, 2006.

CONTENTS

List Of Figures

List Of Tables

Abbreviations

ABR	Auditory brainstem response
ARHL	Age related hearing loss
ATF-1	Activating transcription factor-1
BDNF	Brain derived neurotrophic factor
Cdh23	Cadherin-23
CD:CV	Common disease: common variant hypothesis
CPBP	Core promoter binding protein
DOPAE	Distortion product otoacoustic emission
EGR	Early growth response factor
EMSA	Electrophoretic mobility shift assay
Gfi-1	Growth factor independence-1
HGMP	Human genome mapping project
IHC	Inner hair cell
NCBI	National Centre for Biotechnology Information
NIHL	Noise induced hearing loss
OC-2	organ of Corti-2
OHC	Outer hair cell
PCR	Polymerase chain reaction
PMCA2	Plasma membrane calcium ATPase isoform 2
ROS	Reactive oxygen species
SOD1	Superoxide dismutase-1
SP1	Stimulating protein-1
SSCP	Single strand conformational polymorphism analysis
TUNEL	Terminal dUTP nick end labelling
USH	Usher syndrome

1.0 Introduction.

1.1 An introduction to hearing loss.

Hearing loss is one of the most common, costly and least understood disabilities of man. The most recent figures scored in 2002 estimate that worldwide 250 million people have a disabling hearing impairment (http://www.who.org). In the U.K. alone, approximately 9 million people suffer with hearing loss and of these around 75 % have developed the hearing loss with increasing age (http://www.rnid.org). Hearing loss is often described as a 'hidden' or 'invisible' disability. Indeed, it is not always immediately apparent that a person is deaf or hearing impaired. Yet, this seemingly innocuous, invisible disability has severe consequences. Hearing loss deprives sufferers of their main means of communication and can lead to social isolation, loneliness and depression.

Hearing loss can manifest at any age; congenital deafness (present at birth) is known to affect approximately 1 in 1000 children and a further 1 in 1000 develop permanent hearing loss before adulthood (Morten, 1991; Mason and Herrmann, 1998; Parving, 1999). In addition, many adults go on to develop progressive hearing loss as they age; 16% of adults (17-80 years) have a hearing impairment of at least 25 decibels and for the 60-plus population this figure is more than 30%, increasing to 60% for people aged over 70 years of age (Davis 1989 and 1995). Moreover, the scale of this problem is set to increase as life expectancy increases. This presents a major public health problem and is a significant economic burden. According to the world health organisation adult onset hearing loss ranks 15th under the leading causes of the 'global burden of disease' and 2nd in the leading causes of 'years lived with a disability' (http://www.who.org).

The causes of hearing loss are diverse. Approximately 50% of congenital cases of hearing loss are due to genetic defects (Morton, 1991). Around 25% of cases are due to environmental factors such as bacterial and viral infections, prematurity and exposure to ototoxic agents. Of the remaining 25% of cases the etiology is obscure but a genetic basis is suspected (Morton, 1991; for review see Schrijver 2004). In contrast, late onset hearing loss of sensorineural origin, which is exhibited by a large proportion of the ageing population is thought to be due to a complex genetic – environmental interplay within which acoustic trauma is a major factor (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Ohlemiller, 2004; Gates and Mills, 2005).

Congenital hearing loss attributed to genetic origins can present as an isolated condition termed non-syndromic deafness or it can present in conjunction with another phenotype(s) elsewhere in the body in which case it is termed syndromic deafness. Currently, greater than 400 syndromes have been reported in which hearing loss is a characteristic feature and within which the pathology varies enormously (for reviews see Keats and Berlin, 1999; Friedman et al, 2003). Non-syndromic deafness is the more common of the two, accounting for about 70% of cases (Morten, 1991) and generally is mostly sensorineural in origin (for review see Schrijver, 2004). It is a highly heterogeneous trait, at present nearly 100 loci have been reported from which 39 genes have been identified (Van Camp and Smith, Hereditary Hearing Loss Home Page, http://webhost.ua.ac.be/hhh/) and it can be further categorized according to the mode of inheritance. The majority of genes that cause non-syndromic deafness present as autosomal recessives accounting for around 77% of cases; these loci are denoted DNFB. Autosomal dominant genes account for around 22% of the remaining cases and these loci are denoted DFNA. Only 1% is X-chromosome linked (DFN) and less than 1% show mitochrondrial modes of inheritance (Morton, 1991). Generally, autosomal dominant non-syndromic hearing loss is post-lingual (develops after acquisition of speech) and progressive whereas, autosomal recessive non-syndromic hearing loss is prelingual (develops before acquisition of speech) and is severe to profound (for reviews see Bitner-Glindzicz, 2002; Petersen, 2002; Schrijver, 2004).

The basis of this project is to investigate genetic susceptibility to late onset hearing loss of sensorineural origin, which is exhibited by a large proportion of the ageing population. For the purpose of this thesis the term *late onset hearing loss* was adopted as the nomenclature of choice rather than the terms presbycusis or age-related hearing loss (ARHL) which have frequently been used to describe the common, symmetric, sensorineural high frequency loss of hearing typical with advancing years (Zheng et al, 1998; Alam et al, 2001; Unal et al, 2005; for reviews see Fransen et al, 2003; Gates and Mills, 2005). There are conflicting reports in the literature regarding definition of the term presbyacusis (Zheng et al, 1998; Alam et al, 2001; Seidman et al, 2000; Jennings and Jones, 2001). Whereas, the term ARHL is somewhat misleading it indicates hearing loss solely due to the intrinsic process of auditory ageing. When, in reality ARHL is an extremely complicated trait and many factors are thought to contribute to the pathology. A complex genetic environmental interaction is suspected to underlie the aetiology (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Ohlemiller, 2004; Gates and Mills, 2005) and delineating the effect of individual

exogenous factors such as noise trauma and exposure to ototoxic agents from the ageing process itself is problematic if not impossible. Furthermore, to compound matters accumulating evidence suggests that ARHL and noise-induced hearing loss, NIHL (hearing loss caused by prolonged exposure to moderate level sound) are not separate pathologies but overlapping phenomena (for reviews see Ohlemiller, 2004; Gates and Mills, 2005). Therefore, I have adopted the term *'late onset hearing loss'* as it encompasses the multi-factorial nature of this disease. Defining late onset hearing loss is discussed further in section 1.5.1.

Delineating the genetic basis of late onset hearing loss is extremely important not least because of its significant social and economic burden but also because at present of all common and complex traits the fundamental pathophysiology of late onset hearing loss is probably the least well understood. Unravelling the molecular basis of the hearing process itself has been a slow process especially when compared alongside progress made in other sensory systems such as the visual system; this coupled with the complexity of late onset hearing loss has hindered our understanding of this disease at the molecular level. Significant limiting factors have been the difficulties imposed by the inaccessibility of the inner ear and in obtaining sufficient amounts of cochlea material to study. In addition, the highly heterogenous nature of congenital hearing loss has further compounded the difficulty and hence, slowed progress.

At the onset of this project no common genetic variants had been identified as a risk factor for susceptibility to late onset hearing loss. Understanding the common genetic variants that contribute to late onset hearing loss is extremely important; it will help clarify the underlying disease mechanism and molecular pathways involved. This will help pave way for therapeutic intervention to slow or ultimately prevent loss of hearing with advancing age. It will also complement the knowledge gained from the study of rare monogenic forms of syndromic and non-syndromic hearing loss and extend our understanding of the basic hearing process.

During this introduction the structure and function of the ear will be discussed with emphasis on the cochlea and the sensory hair cells within - highly specialised auditory cells crucial for mechanoelectrical transduction and loss of which is a major cause of late onset hearing loss (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills et al, 2005). The complex transcription factor interactions and Notch signalling events that underlie auditory hair cell development will be summarised followed by a discussion of how elucidation of genes involved in syndromic and non-syndromic forms of hereditary

hearing loss have driven forward tremendous progress in our understanding of the hearing process at the molecular level and how theoretically any of these genes could be involved in susceptibility to late onset hearing loss. The pathophysiology of late onset hearing loss and the cellular and molecular mechanisms that underlie development of this disease will be explored together with methods available to study the genetics of this complex disease. Finally, candidate genes for susceptibility to late onset hearing loss at the onset of this project and that have emerged during the course of this project will be discussed with emphasis on *Brn-3c* (also known as *Brn3.1* and *POU4F3*), the candidate gene for susceptibility to late onset hearing loss that is the focus of this project.

1.2 The structure and function of the mammalian ear: the auditory pathway.

The adult mammalian ear is a highly complex, specialized organ and can be simply divided into three main components: the external, middle and inner ear (see Fig. 1.1). A prominent funnel shaped auricle and the external acoustic meatus form the external ear; sound waves are collected by the former and the latter acts as the auditory canal to conduct sound waves inwards to the tympanic membrane which in response vibrates (Lee, 1999; Moore and Dalley, 1999). Hence, its common name, the eardrum. The tympanic membrane is a thin, almost circular covering that partitions the external and middle ear. Hearing loss manifesting from a defect of the external or middle ear is termed conductive (Lee, 1999). The middle ear resides within the petrous part of the temporal bone and is an air filled chamber termed, the tympanic cavity across which extend three auditory ossicles: the malleus, incus and stapes to form the ossicular chain, see Fig. 1.1 (Lee, 1999; Moore and Dalley, 1999). The lower portion of the malleus embeds into the tympanic membrane and at the far end of the chain the base of the stapes, known as the footplate, occupies the oval window of the inner ear. Fixation of the stapes footplate to the oval window results in otosclerosis, a common form of conductive, adult onset hearing loss treatable by corrective surgery (for review see, Menger and Tange, 2003). The incus resides between the malleus and stapes and articulates with them. The middle ear functions to transform acoustic energy from the medium of air to the medium of fluid. This is achieved by the tympanic membrane and ossicular chain acting in concert; the vibrating movements of the tympanic membrane are transmitted along the ossicular chain to the oval window of the inner ear (Lee, 1999; Moore and Dalley, 1999). Displacement of the ossicular chain is directly related to the frequency and intensity of the sound input and transmission is most efficient at frequencies crucial for understanding speech (500 to 3000 Hz) (Lee, 1999; Moore and Dalley, 1999). The ageing process has minimal affect on the middle ear; the components of the middle ear are subject to age-related changes but this seems to have negligible effect on their function (Wiley et al, 1999).

The inner ear is highly intricate; it has evolved to process sounds across six and nine octaves (for review see Nobili et al, 1998). Continuous with the middle ear in the petrous part of the temporal bone, the inner ear is composed of a very sophisticated closed system of membranous ducts and chambers known as the membranous labyrinth (for review see Forge and Wright, 2002). This endolymph-filled membranous labyrinth

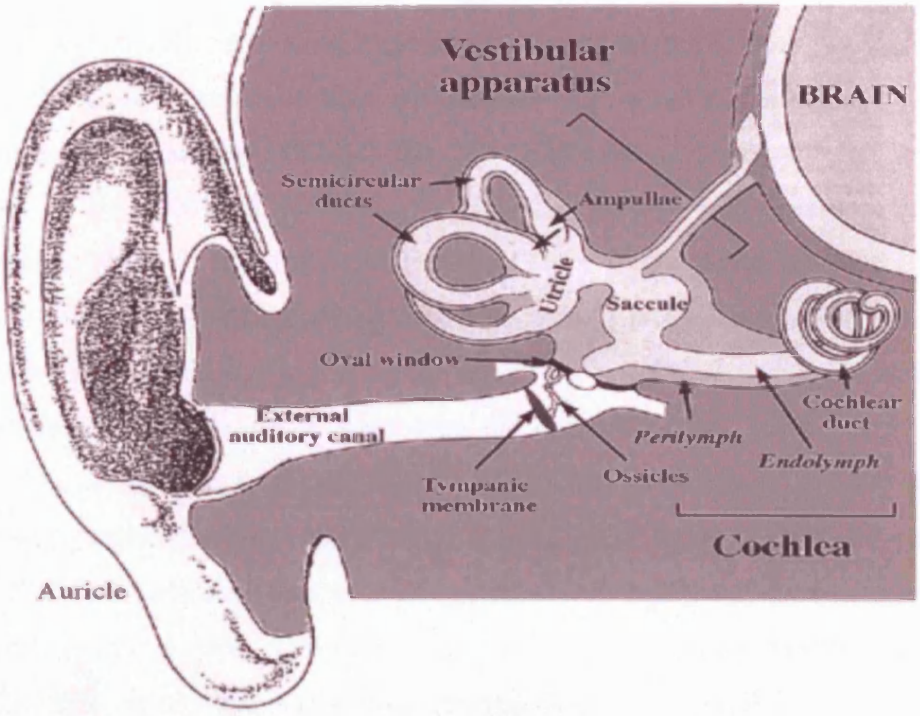


Figure 1.1 A schematic drawing of the mammalian ear. The three main sections of the mammalian ear are shown: the external ear (auricle and external auditory canal), the middle ear (auditory ossicles) and the inner ear (vestibular apparatus and cochlea).
(Reproduced from Kalatzis et al, 1998).

is surrounded by a fluid, perilymph, and lies suspended within the cavities of the bony labyrinth - the walls of which consist of dense bone known as the otic capsule (Lee, 1999; Moore and Dalley, 1999). The membranous labyrinth can be subdivided into two chief sections defined by function: the vestibular complex and the cochlea, which constitute the highly specialized sensory organs of the inner ear essential for balance and hearing, respectively (see Fig. 1.1). Hearing loss manifesting from defects in the cochlea or auditory nerve, the VIII nerve, is termed sensorineural (Lee, 1999). The vestibular complex is composed of the vestibule and (at the posterior end) three semicircular canals lined by semicircular ducts and arranged such that they form three planes in space giving an overall three dimensional structure (Lee, 1999; Moore and Dalley, 1999). The vestibule chamber contains the saccule and utricle and the oval window of the inner ear on its lateral wall. Specialized sensory neuroepithelia within the saccule and utricle (the maculae) detects linear acceleration, and specialized sensory neuroepithelia within the semicircular ducts (the cristae ampullae) detects angular acceleration; this collectively maintains the equilibrium essential for normal balance (Lee, 1999; Moore and Dalley, 1999; for review see Forge and Wright, 2002). A key

component of this specialised sensory neuroepithelia is the hair cell, which will not be discussed further for the vestibular complex but will be described in detail for the cochlea where sound processing is performed (see section 1.2.1).

The cochlea located within the cochlear duct contains the auditory machinery to process the sound signals received from the middle ear. It is a remarkable organ "able to respond to sound-driven vibrations of only +/- 0.3nm – the diameter of an atom" (Hudspeth, 1997). It is a highly active structure and compared to the middle ear it is severely affected by the ageing process (for review see Gates and Mills, 2005). The distinctive coil of the cochlea is due to the spiral canal that originates at the vestibule and winds 2.5 times around a bony core, the modiolus (Lee, 1999; Moore and Dalley, 1999). The cochlea duct, the scala media (filled with endolymph) is securely suspended across the perilymph-filled spiral canal by the spiral ligament and the osseous spiral lamina of the modiolus. This in effect, creates two perilymph-filled channels running horizontally along each side of the cochlea duct: the scala vestibuli and the scala tympani. The endolymph has a high potassium ion but low sodium ion concentration whereas the perilymph has a high sodium ion concentration and is low in potassium ions (for review see Forge and Wright, 2002). Maintenance of these differing cochlea fluids is crucial for the hearing process; this is achieved in part by ionic pumps in the stria vascularis located on the lateral wall of the cochlea duct that pump potassium ions into the endolymph (Takeuchi et al, 2000). Hence, ensuring that a high positive resting potential, the endocochlea potential is maintained to drive mechanoelectrical transduction for hearing.

Two membranes, the vestibular and basilar membranes form the upper and lower surfaces of the cochlea duct, respectively. Along the length of the basilar membrane lies the organ of Corti and it is here that the specialized sensory neuroepithelia of the cochlea is located and mechanoelectrical transduction mediated by auditory hair cells is performed (see Fig.1.2). The organ of Corti is a highly differentiated and ordered structure consisting of two morphologically distinct sets of hair cells: inner hair cells (IHCs) and outer hair cells (OHCs) in addition to an array of supporting cells: Deiter's cells, Hensen's cells and pillar cells (for review see Forge and Wright, 2002). Typically, three rows of OHCs are precisely arranged along the longitudinal axis of the organ of Corti clearly distinguishable from a single row of IHCs (for review see Forge and Wright, 2002). A gelatinous membrane termed, the tectorial membrane overlies the hair cells. The sensory hair cells do not directly contact each other, but instead are surrounded by supporting cells (see Fig. 1.2). This alternating hair

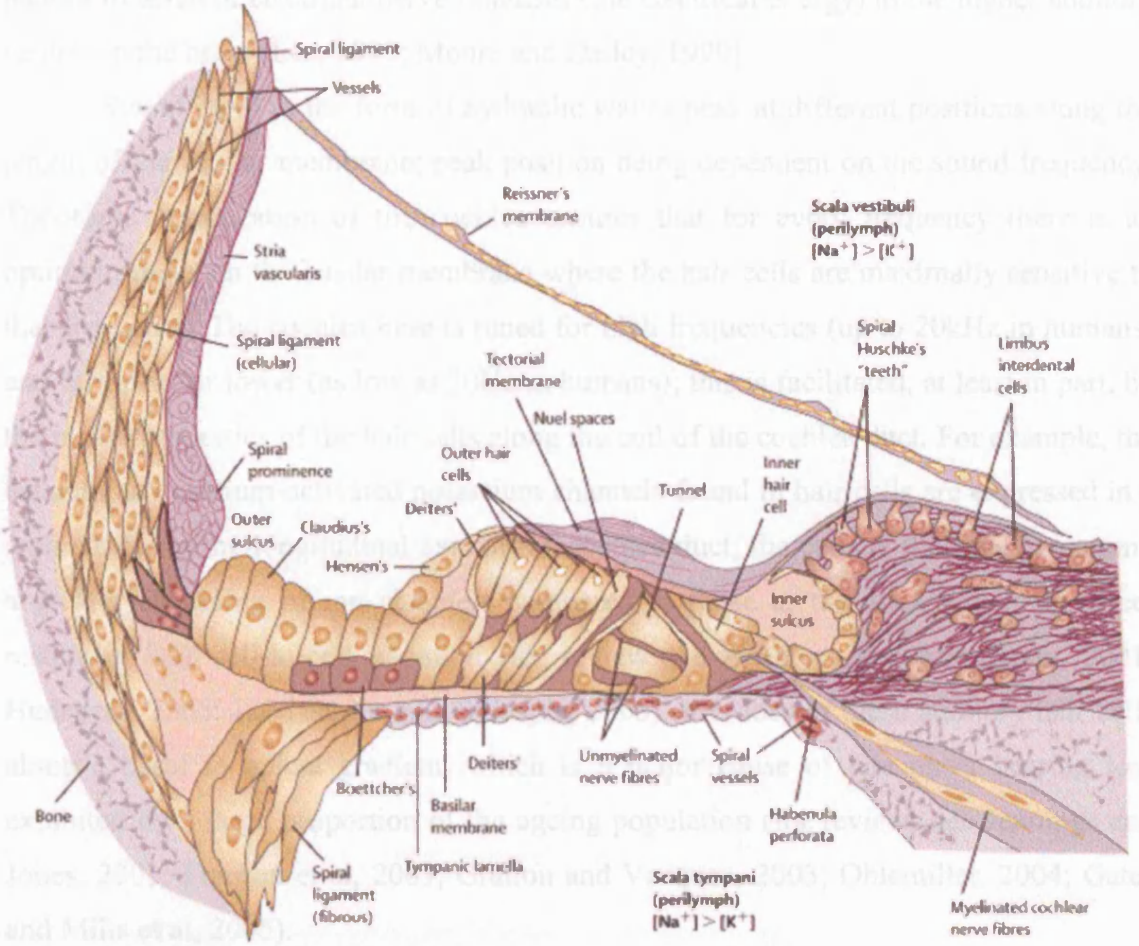


Figure 1.2 A cross-section of the cochlea to show the sensory neuroepithelia for hearing. Inner and outer hair cells of the organ of Corti are shown surrounded by supporting cells (Deiter's cells and Hensen's cells). The organ of Corti lies on the basilar membrane that extends along the coil of the cochlea duct. The tectorial membrane overlies the hair cells.
(Reproduced from Gates and Mills, 2005).

cell: supporting cell arrangement forms the predominant cell network in the organ of Corti.

In the cochlea, the sound input is processed from mechanical energy into hydraulic energy into its final form, electrical energy. This is achieved by the vibrations (the mechanical energy) of the footplate of the stapes in and out of the oval window creating waves of hydraulic pressure (the hydraulic energy) in the perilymph of the vestibule (Lee, 1999; Moore and Dalley, 1999). The hydraulic waves ascend towards the apex of the cochlea via the scala vestibuli. In doing so the force of the hydraulic wave displaces the basilar membrane this in turn triggers mechanoelectrical transduction mediated by the sensory hair cells (see section 1.2.1 below) and transmission of a complex array of frequencies that constitutes the sound input as a

pattern of afferent electrical nerve impulses (the electrical energy) to the higher auditory centres in the brain (Lee, 1999; Moore and Dalley, 1999).

Sound input in the form of hydraulic waves peak at different positions along the length of the basilar membrane; peak position being dependent on the sound frequency. Tonotopic organization of the cochlea ensures that for every frequency there is an optimum locus on the basilar membrane where the hair cells are maximally sensitive to that frequency. The cochlea base is tuned for high frequencies (up to 20kHz in humans) and the apex for lower (as low as 20Hz in humans); this is facilitated, at least in part, by the innate properties of the hair cells along the coil of the cochlea duct. For example, the calcium and calcium-activated potassium channels found in hair cells are expressed in a gradient along the longitudinal axis of the cochlea duct; the cells at the base containing a greater clustering of ion channels compared to those at the apex which, in effect results in hair cell-to-cell variation (for review see Ashmore and Mammano, 2001; Hudspeth, 2005; LeMasurier and Gillespie, 2005). It is loss of these auditory hair cells along a basal to apical gradient, which is a major cause of late onset hearing loss exhibited by a large proportion of the ageing population (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills et al, 2005).

1.2.1 Hair cells: the mechanoelectrical transducers of the cochlea.

Hair cells are crucial for hearing; they are the mechano-electricial transducers of the cochlea. Loss of these cells from the cochlea is a major cause of late onset hearing loss (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills et al, 2005) and abnormalities in these cells are responsible for various forms of hereditary non-syndromic congenital deafness (see section 1.4 and Van Camp and Smith, Hereditary Hearing Loss Home Page, http://webhost.ua.ac.be/hhh/). Paramount to their function is their apical specialization for the detection of movement, the mechanosensory hair bundle. Composed of around 20-300 specialized microvilli, termed stereocilia the hair bundle protrudes towards the overlying tectorial membrane within the cochlea duct. The stereocilia are composed of a rigid core of parallel actin filaments that gradually narrows as it anchors the stereocilia securely into the cuticular plate, a solid support composed of actin located on the apical surface of hair cells (for review see Steel and Kros, 2001; Forge and Wright, 2002). Three-dimensionally the stereociliary hair bundle resembles a pyramid-shaped array

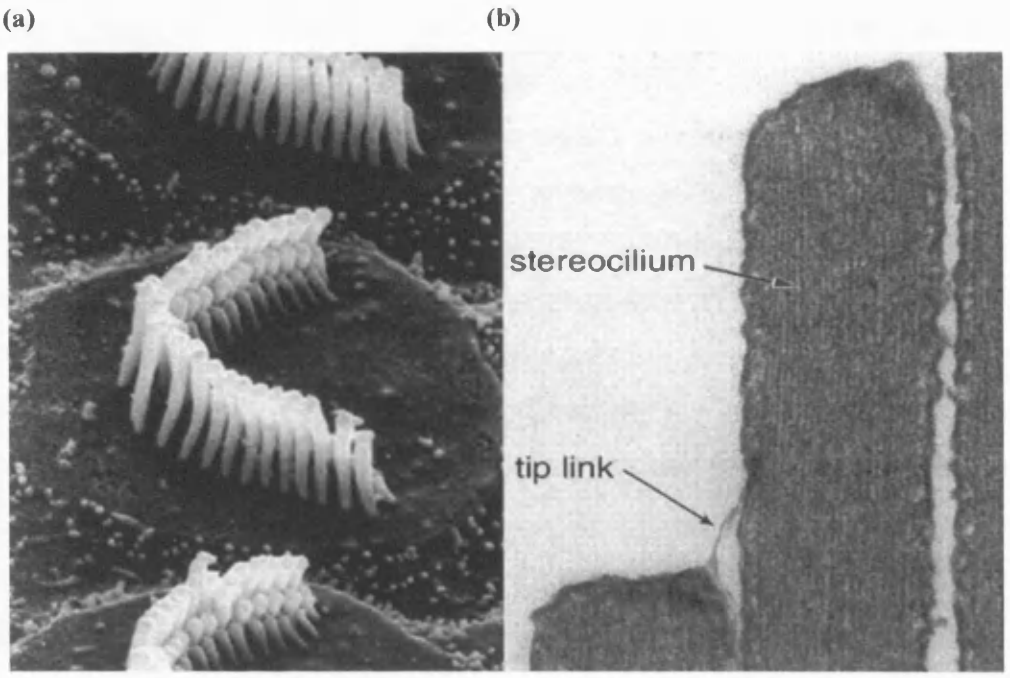


Figure 1.3 Hair cell stereocilia and their tip-links. **(a)** Scanning electron microscopy of the stereociliary hair bundle of an outer hair cell; three-dimensionally this resembles a pyramid-shaped array. **(b)** Fine detail of the tip-link; crucial stereocilium – stereocilium connections that link each stereocilium to the side of their immediate taller neighbour.
(Reproduced from Petit et al, 2001).

(see Fig. 1.3). Along the length of the stereocilia extend extracellular lateral links that formstereocilium - stereocilium connections within the array; lateral links are thought to be important in hair bundle stability (for review see Forge and Wright, 2002). At the tip of each stereocilium extend stereocilium - stereocilium connections termed tip-links that connect each stereocilium to the side of their immediate taller neighbour, see Fig. 1.3 (Pickles et al, 1984, Kachar et al, 2000); tip-links are thought to be crucial for mechanoelectrical transduction. It has been proposed that Cadherin-23 (Cdh23) is a major component of the tip-link (Siemens et al, 2004; Sollner et al, 2004; Tsuprun et al, 2004); Cdh23 is thought to link stereociliary tips through a homophilic binding mechanism (Siemens et al, 2004). However, this evidence is based on expression studies in mice at early embryonic stages and has been contradicted by others who report that Cdh23 is absent from the apex of stereocilia in adult mice (Lagziel et al, 2005; Michel et al, 2005). It has been proposed that the mechanoelectrical transduction channels are attached to the ends of the tip links (for review see Hudspeth, 2005).

Displacement of the basilar membrane deflects the hair bundle resulting in a 'shearing' motion of the stereocilia against the mass of the overlying tectorial

membrane. This causes the stereocilia to pivot towards their nearest taller neighbour creating a wave of movement within the pyramid-shaped stereociliary hair bundle (for review see Hudspeth, 2005). It has been proposed that the deflection of the stereocilia causes the tip-links to act as 'gating springs'; an increase in tension opens the mechanoelectrical transduction channels causing an influx of cations – probably predominantly potassium and to less extent, calcium into the hair cell depolarizing it (Corey and Hudspeth, 1983; Howard and Hudspeth, 1988; Holt and Corey, 2000). However, other evidence is not in agreement with this proposal; structural examination of the tip link at high resolution suggests that the tip link itself, is not the gating spring (Kachar et al, 2000). At present, the molecular identity of the mechanoelectrical transduction channel is unknown. A member of the transient receptor potential (TRP) superfamily of ion channel proteins, TRPA1, has been suggested as a strong candidate based on a repertoire of data by Corey et al, 2004. However, recent evidence disputes this; mice knockout for the TRPA1 gene are not deaf and exhibit a normal auditory brainstem response (Kwan et al, 2006).

Depolarization of OHCs and IHCs in this manner leads to defined cellular responses reflective of their distinct morphological nature. The cylindrical shaped OHCs receive the majority (about 80%) of the efferent innervation of the cochlea and are responsible for 'cochlea amplification' a process that augments the motion of the basilar membrane so that the sound input is sensitised 100-fold (for reviews see, Nobili et al, 1998; Ashmore and Mammano, 2001; Hudspeth, 2005; LeMasurier and Gillespie, 2005). Structurally, the arrangement of the OHCs facilitates this; their stereociliary hair bundle contacts the overlying tectorial membrane and they ride on the motion of the basilar membrane in such a way that minute forces are channelled back into the dynamics of this membrane (for reviews see Nobili et al, 1998; Ashmore and Mammano, 2001; Hudspeth, 2005; LeMasurier and Gillespie, 2005). How this mechanism of amplification occurs has yet to be clarified, but one likely proposal is the mechanism of 'somatic electromotility'. This features a cellular motor that is known to pack densely in the lateral plasma membrane of the OHCs. In this model a change in the endocochlea potential resulting from the shearing motion of the stereocilia drives a conformational change in the motor molecule that causes the OHCs to characteristically lengthen and shorten thus forming an active feedback mechanism whereby the mechanics of the basilar membrane is enhanced (for review see Nobili et al, 1998; Ashmore and Mammano, 2001; Hudspeth, 2005; LeMasurier and Gillespie, 2005). Prestin, a member of the solute carrier family 26 (SLC26A) of anion transporters has

been confirmed as the motor molecule (Zheng et al, 2000a, Belyantseva et al, 2000, Liberman et al, 2002; for review see Dallos and Fakler, 2002).

In comparison, the flask-shaped IHCs receive the majority of afferent innervation of the cochlea (around 90-95%) and rapidly transmit the sound signal (fine-tuned and maximally amplified by the OHCs) to the higher auditory centres in the brain. IHCs lack axons and dendrites; at their base they make afferent synaptic contacts with the axonal terminals of the cochlea branch of the VIIIth nerve (for review see Hudspeth, 2005). Thus, it is IHCs that are the primary sensory receptors of the cochlea. This role is facilitated by unique L-type calcium channels at their basolateral surface; voltage-gated calcium channels that activate extremely rapidly but inactive very slowly (Platzer et al, 2000). Depolarisation of IHCs drives the opening of the voltage-gated L-type calcium channels leading to a rise in intracellular calcium in the IHC cytoplasm. This in turn stimulates neurotransmitter release at glutamatergic synapses on the auditory afferent nerves for signal propagation and hence, timely couples IHC depolarisation to neurotransmitter release for hearing (for reviews see Hudspeth, 2005; LeMasurier and Gillespie, 2005).

1.2.2 The vulnerability of auditory hair cells.

Clearly, hair cells of the cochlea are highly specialised sensory cells; they play a paramount role in the hearing process. Loss of IHCs causes severe deafness. Whereas, loss of functional OHCs but not IHCs causes a shift in hearing thresholds (around 60-80 decibels) this abolishes the fine-tuning of the sound input, which is crucial for frequency discrimination (for review see Forge and Wright, 2002; Hudspeth, 2005). Thus, perhaps not surprising it is loss of auditory hair cells that is a major cause of late onset hearing loss (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills et al, 2005).

Loss of auditory hair cells concomitant with loss of hearing is exacerbated by the fact that in mammal's hair cell loss cannot be replaced by mitosis; auditory hair cells are terminally differentiated. Furthermore, mammals do not posses the inherent ability to regenerate cochlea hair cells from surrounding supporting cells although this has been documented in other species including fishes, amphibians, and birds (for reviews see Feghali et al, 1998; Staecker and Van De Water, 1998; Bermingham-McDonogh and Rubel, 2003; Matsui et al, 2005). At birth we have a limited number of hair cells, around 16,000 are present per cochlea (for review see Hudspeth, 2005) and these must

be maintained in our lifespan if we are to retain our ability to hear into old age. However, the intrinsic propensity of auditory hair cells to be maintained is exacerbated by the inherent nature of these cells; they are extremely delicate and tremendously susceptible to damage by exogenous stress. Many environmental factors are thought to accelerate development of late onset hearing loss via their effects on hair cells. Acoustic trauma is probably the most widely recognised environmental stress to auditory hair cells (Emmerich et al, 2000; Chen and Fechter, 2003; for review see Lynch and Kil, 2005). This is perhaps not surprising given that the delicate stereociliary hair bundles on the apical surface of the hair cells intercept sound-induced vibrations from the middle ear. Acoustic trauma can manifest from exposure to an isolated incident of intense sound or in comparison, and perhaps rather simply, the cumulative hub of daily noise exposure – a factor that has the potential to vary widely across different individuals over their lifespan. Viral infections and ototoxic therapeutic agents such as cisplatin used in chemotherapy, and gentamicin an aminoglycoside antibiotic that is widely used as an antimicrobial agent are additional sources of exogenous stress that are extremely deleterious to hair cells (for review see Rybak and Whitworth, 2005). Certainly, for some people development of hearing loss induced by exposure to aminoglycoside antibiotics is due to presence of a predisposing mutation in a mitocochondrial gene. Three mutations: A1555G, C1494T and delT961Cn in the mitochondrial 12S ribosomal RNA gene all seem to confer increased susceptibility to the ototoxic effects of aminoglycoside antibiotics (for review see Fischel-Ghodsian, 2003; Fischel-Ghodsian et al, 2004).

Studies with animals show that many acquired environmental stresses including acoustic trauma and ototoxic agents cause hair cell death via apoptosis (for reviews see Lefebvre et al, 2002; Holley et al, 2005; Cheng, et al, 2005; Rybak and Whitworth, 2005). In addition, whether exogenous stress is inflicted through use of aminoglycoside antibiotics, cisplatin or through exposure to acoustic trauma, the outer hair cells appear the most susceptible to damage, particularly in the basal region of the cochlea (for reviews see Lynch and Kil, 2005; Rybak and Whitworth, 2005). This is evident in late onset hearing loss, which is characterised by hearing loss that predominates in the high frequencies (for reviews Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills et al, 2005).

Maintenance of hearing into old age is thought to be strongly dependent upon the continued function and expression of genes that are involved in the repair and maintenance of auditory hair cells in response to exogenous stress and that are needed

for long-term survival. Hence, common sequence variants in the general population that impair the functioning or expression of such genes could be a risk factor for susceptibility to late onset hearing loss. Brn-3c is the candidate gene for susceptibility to late onset hearing loss that is the focus of this project and appears to function as a long-term pro-survival factor for auditory hair cells (Erkman et al, 1996, Xiang et al, 1997, Xiang et al, 1998 and Vahava et al, 1998); this will be discussed further in sections 1.8 and 1.9.

1.3 Development of sensory hair cells.

The mammalian inner ear originates from the otic placode, a thickening of the ectoderm adjacent to the dorsal hindbrain that also gives rise to the VIIIth ganglion neurons that will innervate the sensory hair cells of the inner ear (for reviews see Fritzsch and Beisel, 2001; Fekete and Wu, 2002; Barald and Kelley, 2004). During embyrogenesis the octic placode invaginates to form the otocyst, a simple epithelial sac. The otocyst develops via a complex pattern of morphogenetic events involving proliferation, apoptosis, migration and differentiation to form the membraneous labyrinth of the inner ear; an elaborate fluid filled structure that contains the specialised sensory neuroepithelia for hearing, motion and gravity (for reviews see Fritzsch and Beisel, 2001; Fekete and Wu, 2002; Barald and Kelley, 2004). During this process various transcription factor interactions underlie the complex cell fate decisions that are made to ensure specification of various cell types that shape the intercalated sensory compartments that form the inner ear (for reviews see Fekete and Wu, 2002, Barald and Kelley, 2004).

The sensory neuroepithelia for hearing resides within the organ of Corti that extends along the coil of the cochlea duct. As discussed, the organ of Corti is a highly differentiated and ordered structure consisting of an alternating arrangement of hair cells surrounded by supporting cells. In most mammals, a single row of inner hair cells is orderly arranged along the coil of the cochlea duct flanked by three rows of outer hair cells on the outer side of the spiral (for review see Forge and Wright, 2002). The sequence of developmental events that dictates this specific cellular organization is known to act along a basal to apical developmental gradient and appears to be based around a developmental model known as lateral inhibition (for reviews see Bryant et al, 2002; Fekete and Wu, 2002; Hawkins and Lovett, 2004). Understanding the developmental pathway that underlies auditory hair cell development is extremely important; auditory hair cells have not been shown to regenerate in mammals and by understanding how these hair cells are specified future therapeutic approaches to hearing loss involving gene and stem cell therapies will be facilitated. Some of the main aspects of auditory hair cell development will be discussed.

1.3.1. Notch signalling in hair cell development in the auditory system.

Lateral inhibition is a mechanism by which cells of a common origin go on to adopt different fates. Simplistically, the basis of this model is that from an equivalent

population of precursor cells once a cell is committed to its primary fate this inhibits surrounding cells from committing to the same fate. Consequently, the surrounding cells follow an alternative developmental pathway. In the developing organ of Corti lateral inhibition appears to be mediated by Notch signalling (for reviews see Bryant et al, 2002; Fekete and Wu, 2002; Hawkins and Lovett, 2004). Notch proteins are trans-membrane receptors that are activated by direct cell-cell interactions with a ligand-expressing cell; they are important components of intracellular signal transduction cascades that regulate specification of cell fate (for review see Weinmaster, 1998). Upon ligand binding it is proposed that proteolytic cleavage results in the dissociation of the intracellular domain of notch, which then translocates to the nucleus to activate target genes (for review see Weinmaster, 1998).

Zine et al, 2000 provided clear evidence that Notch signalling is important in determining whether cells differentiate as hair cells or supporting cells. An investigation into the spatiotemporal expression pattern of Notch1 receptor and its ligands (Jag1 and Jag2) in the developing auditory sensory neuroepithelia of the rat revealed an overlapping but distinct expression pattern. During early hair cell differentiation (between E18 and birth) Notch 1 and Jag 1 expression predominate in supporting cells whereas *Jag 2* predominates in newly formed hair cells. Shortly after birth the expression dynamics change, hair cells as well as supporting cells express Notch 1 and Jag 2 expression is down-regulated in hair cells. The significance of this expression pattern on development of the orderly hair cell-supporting cell network is evident by treatment of developing organ of Corti explants in the rat with antisense oligonucleotides directed against either Notch1 or Jag1. Interference of Notch signalling in this manner results in disruption of the normal cellular organization in the organ of Corti; extra rows of hair cells are generated that are often not separated by supporting cells. In agreement with this, targeted deletion of the Jag2 gene in mice leads to an extra row of inner and outer hair cells presumably through disruption of Notch signalling to limit hair cell formation (Lanford et al, 1999). Similarly, Delta1 is another Notch ligand that is expressed by hair cells of the inner ear (Morrison et al, 1999) and in zebrafish, the deltaA^{dx2} allele, a dominant negative mutant of the zebrafish deltaA gene leads to an increase in the number of hair cells in the zebrafish inner ear concomitant with a decrease in the size of the supporting cell population (Riley et al, 1999). In support of this data, recently it has been shown that targeted deletion of the delta1 gene in mice leads to excess production of auditory hair cells (Brooker et al, 2006). Collectively this evidence is consistent with the concept whereby hair cells use Notch signalling to

ensure that adjacent cells adopt different cell fates and hence, the differentiation of neighbouring cells into supporting cells. More recent evidence indicates that lateral inhibition of hair cell differentiation may not be the only role of Notch in inner ear development. Over-expression of activated Notch in the chick otocyst reveals that Notch also appears to function in specification of the prosensory patch from which hair cells and supporting cells will subsequently arise (Daudet and Lewis, 2005).

1.3.2 Transcription factors involved in hair cell development in the auditory system.

Development of the orderly hair cell-supporting cell network in the organ of Corti is not just dependent on Notch signalling. Many transcription factors are also involved in this process and have been established to play crucial roles in hair cell specification, differentiation and maintenance. In addition, it appears that at least some of these transcription factors exhibit interplay with components of the Notch pathway (for reviews see Bryant et al, 2002; Fekete and Wu, 2002; Hawkins and Lovett, 2004).

Math1 encodes a basic helix-loop-helix transcription factor. It is a mouse homologue of the *Drosophila melanogaster* proneural gene atonal and is crucial for the generation of sensory hair cells in mice (Bermingham et al, 1999; Zheng and Gao, 2000). Bermingham et al, 1999 created Math1-null mice by replacing the coding region of Math1 with β-galactosidase (β-gal). Examination of β-gal expression in Math1-null (β-gal/β-gal) and heterozygous (β-gal/-) mice revealed that β-gal staining recapitulated the endogenous Math1 expression pattern. Math1 expression (β-gal) is detected in the prospective sensory epithelia of the otic vesicle at E12.5 before the hair cells start to differentiate. Expression continues throughout the epithelia from which the sensory organs will arise and in (β-gal/-) mice is restricted to the hair cells by E18.5. Scanning electron and transmission electron microscope analysis showed that these Math1-null (β-gal/β-gal) mice failed to form hair cells but not supporting cells. Progenitor cells within the sensory epithelia appear unable to differentiate into hair cells; hair cell specific markers fail to stain in Math1-null mice at a time by which the markers are expressed in wild type mice. This knockout approach indicates that Math1 is required for the specification of hair cells. A Math1 transgenic approach in postnatal rat cochlear explant cultures is in agreement with this (Zheng and Gao, 2000). The overexpression of Math1 in the non-sensory cells of the greater epithelial ridge (adjacent to the organ of Corti) where it is not normally expressed is sufficient to drive them into hair cells. In

addition, overexpression of Math1 in supporting cells of the utricle, part of the vestibular complex induces their trans-differentiation into hair cells (Zheng and Gao, 2000). Clearly Math1 is an important developmental gene involved in hair cell fate determination.

Recent evidence suggests that although Math1 expression is crucial for the selection and/or differentiation of hair cells, it does not underlie specification of the initial cell lineage that forms the common origin from which sensory hair cells and supporting cells will arise. Instead, Math1 appears to act downstream of this cell fate decision by selecting cells from this already specified domain, a patch of cells named the sensory primordium. This is an area of cells that have undergone their final mitosis, from which the complete population of sensory hair cells and supporting cells that form the mature organ of Corti will originate. Evidence suggests expression of Math1 in these cells once they have exited the cell cycle is the trigger to become hair cells rather than supporting cells. It is proposed that lateral inhibition restricts Math1 expression in the surrounding cells, which go on to develop as supporting cells (Chen et al, 2002).

More recently, a transgenic approach using an adenoviral vector has shown that Math1 transfected into the cochlea of adult guinea pigs profoundly deafened by systemic administration of kanamycin induces generation of hair cells and improves hearing thresholds suggesting regeneration of functional hair cells (Izumikawa et al, 2005). Direct transdifferentiation of existing supporting cells into hair cells is thought to be the mechanism behind the hair cell regeneration. This data holds promise; it shows that manipulating expression of an important developmental gene is a potential strategy for gene therapy in the mature cochlea. However, the data presented by Izumikawa et al is in its infancy. Although regenerated IHCs seem normal morphologically regenerated OHCs do not appear fully differentiated and acquisition of their normal functional role to fine-tune and amplify the sound input seems unlikely.

Math1 is clearly a major positive regulator of hair cell specification. Hes-1 and Hes-5 are two basic helix-loop-helix transcription factors that participate in development of the orderly hair cell-supporting cell network in the organ of Corti by functioning as negative regulators of inner and outer hair cell differentiation, respectively (Zheng et al, 2000b; Zine et al, 2001). Hes-1 knockout and Hes-5 knockout mice exhibit a significant increase in the number of IHCs and OHCs, respectively (Zheng et al, 2000b; Zine et al, 2001). The expression of Math1 in these supernumerary hair cells (Zine et al, 2001) coupled with the fact Hes-1 overexpression in postnatal rat cochlea explant cultures can prevent production of supernumerary hair cells induced by

ectopic Math1 expression (Zheng et al, 2000b) has lead to the suggestion that Hes-1 and Hes-5 may regulate hair cell differentiation by antagonising the actions of Math1. Although, whether Hes-1 and Hes-5 directly regulate Math1 expression to limit the number of hair cells that form is not completely clear at present. In neuronal systems there is evidence to implicate that Hes-1 and Hes-5 are regulated by Notch signalling (Ohtsuka et al, 1999). It is possible that this also occurs in the developing sensory epithelia of the cochlea; expression of components of the Notch signalling pathway including Jag2 and Notch1 are modified in the developing organ of Corti of Hes-1 and Hes-5 knockout mice (Zine and Ribaupierre, 2002).

In summary, the current evidence suggests that the decision as to whether a cell that has exited the cell cycle and formed part of the sensory primordium becomes a hair cell or supporting cell depends on a complex interplay between notch signaling, and the opposing actions of positive and negative basic helix-loop-helix transcription factors. Zine and Ribaupierre, 2002 present a good overview of this pathway and this is summarized in Fig.1.4. After specification of the hair cell fate expression of Brn-3c, a POU domain transcription factor is crucial for the mature differentiation of hair cells; targeted deletion of the Brn-3c gene in mice leads to generation of immature hair cells that cannot develop stereociliary bundles (Erkman et al, 1996, Xiang et al, 1997, 1998). Therefore, Brn-3c appears to act downstream of Math1 in hair cell development. In agreement with this, the ectopic overexpression of Brn-3c in postnatal rat cochlear explant cultures does not lead to the generation of supernumerary hair cells (Zheng and Gao, 2000). Brn-3c is the candidate gene for susceptibility to late onset hearing loss that is the focus of this project and will be discussed further in section 1.8 and 1.9.

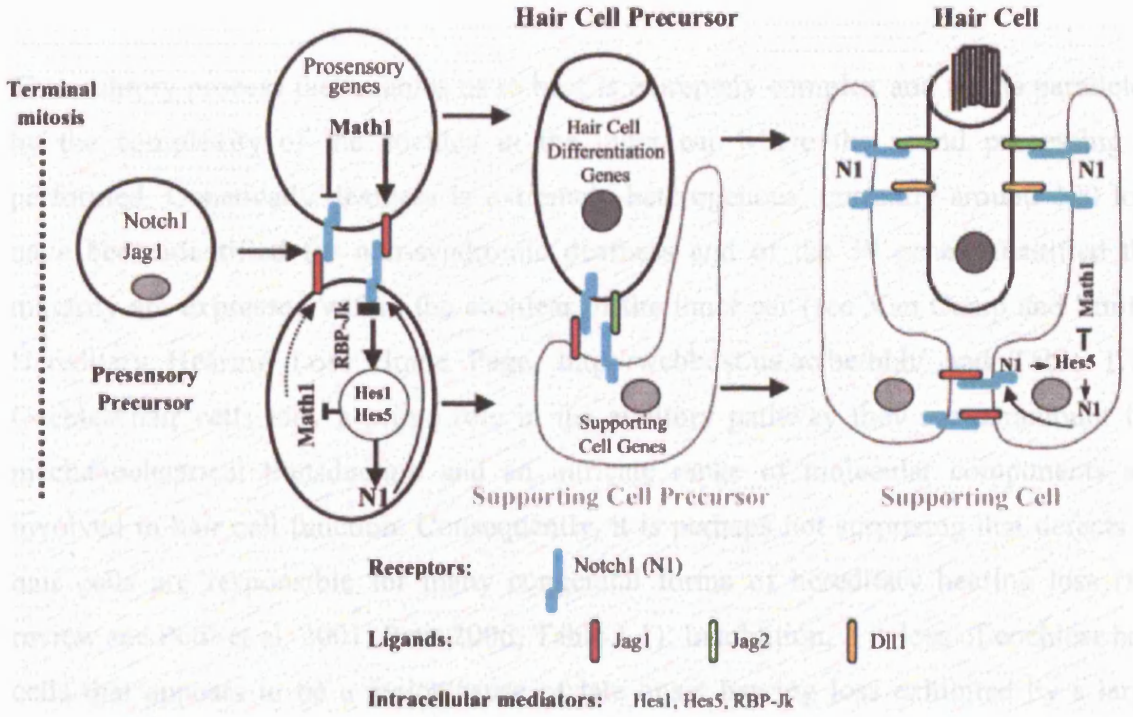


Figure 1.4 A schematic representation of a putative model to summarise the hair cell : supporting cell developmental pathway. The decision as to whether a cell becomes a hair cell or supporting cell depends on a complex interplay between basic helix-loop-helix transcription factors (Hes1 and Hes5) and components of the Notch signalling cascade: the Notch 1 receptor (N1) and the Notch ligands Jag1, Jag2, and Delta1 (Dll1 in diagram). RBP-Jk: recombination sequence specific binding protein at the site Jk (also known as CBF1, core binding factor-1). RBP-Jk is a DNA binding protein that is reported to form a complex with the cleaved intracellular domain of the Notch 1 receptor which then translocates to the nucleus to activate expression of target genes: shown as Hes 1 and Hes 5 in the diagram. Reproduced from Zine and Ribaupierre, 2002.

1.4 The genetics of deafness.

The auditory process that enables us to hear is extremely complex and this is paralleled by the complexity of the cochlea in the inner ear where the sound processing is performed. Genetically deafness is extremely heterogenous; currently around 100 loci have been identified for non-syndromic deafness and of the 39 genes identified the majority are expressed within the cochlear of the inner ear (see Van Camp and Smith, Hereditary Hearing Loss Home Page, http://webhost.ua.ac.be/hhh/ and Table 1.1). Cochlear hair cells play a prime role in the auditory pathway they are paramount for mechanoelectrical transduction and an intricate range of molecular components are involved in hair cell function. Consequently, it is perhaps not surprising that defects in hair cells are responsible for many congenital forms of hereditary hearing loss (for review see Petit et al, 2001; Petit 2006; Table 1.1). In addition, it is loss of cochlear hair cells that appears to be a major cause of late onset hearing loss exhibited by a large proportion of the ageing population (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills et al, 2005). However, in comparison to many congenital forms of hereditary hearing loss the genetics of late onset hearing loss is poorly understood at present.

The mutations responsible for monogenic forms of hereditary hearing loss often have a profound effect on protein function and within the last decade have driven forward great progress in our understanding of the basic hearing mechanism by identification of many key players in the hearing process (for reviews see Steel and Kros, 2001; Bitner-Glindzicz, 2002; Petit 2006). The discovery of novel congenital deafness genes in humans has been facilitated, in part, due to the closeness in homology between the human and mouse genomes. The identification of many *spontaneous* mouse mutants that underlie deafness in mice and subsequent mapping of the genes responsible has aided identification of human deafness genes (for review see Probst and Camper, 1999, Kiernan and Steel, 2000; Ahituv and Avraham, 2002; Avraham, 2003). Similarly, the large-scale mouse mutagenesis screens using the chemical mutagen N-ethyl-N-nitrosourea (ENU) *to induce* random point mutations into the DNA sequence has aided identification of novel genes involved in the hearing process in recent years and many more are still anticipated (for review see Parkinson and Brown, 2002). A thorough discussion of the role of all 39 genes

Table 1.1: Genes underlying hereditary non-syndromic deafness and corresponding mouse mutants (reproduced from Petit, 2005).

Primary defect	Gene	Gene product	Forms of human deafness	Mouse mutants
Hair cells	*MYO7A*	Myosin VIIA (motor protein)	DFNB2 ± retinopathy[a] (Usher 1B) DFNA11	Shaker-1 (*Sh1*)
	MYO15	Myosin XV (motor protein)	DFNB3	Shaker 2 (*Sh2*)
	MYO6	Myosin VI (motor protein)	DFNA22 ± cardiomyopathy[a] DFNB37	Snell's waltzer (*Sv*)
	MYO3A	Myosin IIIA (motor protein)	DFNB30	
	MYO1A	Myosin IA (motor protein)	DFNA48	
	ACTG1	γ-Actin (cytoskeletal protein)	DFNA20 (DFNA26)	
	USH1C	Harmonin (PDZ domain-containing protein)	DFNB18 ± retinopathy[a] (Usher 1C)	Deaf circler (*Dfcr*)
				Deaf circler 2 Jackson (*Dfcr-2J*)
	WHRN	Whirlin (PDZ domain-containing protein)	DFNB31	Whirler (*Wi*)
	CDH23	Cadherin-23 (cell-adhesion protein)	DFNB12 ± retinopathy[a] (Usher 1D)	Waltzer (*V*)
	PCDH15	Protocadherin-15 (cell-adhesion protein)	DFNB23 ± retinopathy[a] (Usher 1F)	Ames waltzer (*Av*)
	TMIE	TMIE (transmembrane domain-containing protein)	DFNB6	Spinner (*Sr*)
	STRC	Stereocilin	DFNB16	
	SLC26A5	Prestin (anion transporter)	DFNB61	*Slc26a5*$^{-/-}$[c]
	ESPN	Espin (actin-bundling protein)	DFNB36, DFNA,[b]	Jerker (*Je*)
	KCNQ4	KCNQ4 (K^+ channel subunit)	DFNA2	
	TMC1	TMC1 (transmembrane channel-like protein)	DFNB7 (DFNB11), DFNA36	Deafness (*Dn*) Beethoven (*Btn*)
	OTOF	Otoferlin (putative vesicle traffic protein)	DFNB9	
	POU4F3	POU4F3 (transcription factor)	DFNA15	*Brn3c*$^{-/-}$[c] Dreidl (*Ddl*)
Non-sensory cells	*GJB2*	Connexin-26 (gap junction protein)	DFNB1, DFNA3 ± keratodermia[a] (Vohwinkel, palmoplantar keratoderma, KID, Bart-Pumphrey)	*Cx26*$^{Drop\ Cre/+}$
	GJB6	Connexin-30 (gap junction protein)	DFNB1, DFNA3 ± keratodermia[a] (KID)	*Cx30*$^{-/-}$[c]
				Cx26$^{+/-}$/*Cx30*$^{+/-}$[c]
	GJB3	Connexin-31 (gap junction protein)	DFNA,[b]	
	SLC26A4	Pendrin (I^--Cl^- transporter)	DFNB4 ± thyroid goiter[a] (Pendred)	*Pds*$^{-/-}$[c]
	CRYM	μ-Cristallin (thyroid hormone-binding protein?)	DFNA,[b]	
	OTOA	Otoancorin (cell-surface protein)	DFNB22	
	CLDN14	Claudin-14 (tight-junction protein)	DFNB29	*Cldn14*$^{-/-}$[c]
	COCH	Cochlin (extracellular matrix component)	DFNA9	
	TMPRSS3	TMPRSS3 (transmembrane serine protease)	DFNB8 (DFNB10)	
	MYH9	Myosin IIA (motor protein)	DFNA17 ± giant platelets[a] (Fechtner)	
	MYH14	Myosin IIC (motor protein)	DFNA4	
	EYA4	EYA4 (transcriptional coactivator)	DFNA10	
	POU3F4	POU3F4 (transcription factor)	DFN3	Sex-linked fidget (*slf*) *Brn4*$^{-/-}$[c]
Tectorial membrane	*COL11A2*	Collagen XI (α2-chain) (extracellular matrix component)	DFNA13 ± osteochondrodysplasia[a] (Stickler 2)	*Col11a2*$^{-/-}$[c]
	TECTA	α-Tectorin (extracellular matrix component)	DFNA8 (DFNA12), DFNB21	*Tecta*$^{-/-}$[c]
Unknown	*HDIA1*	Diaphanous-1 (cytoskeleton regulatory protein)	DFNA1	
	DFNA5	Unidentified	DFNA5	*Dfna5*$^{-/-}$[c]
	WFS1	Wolframin (endoplasmic-reticulum membrane protein)	DFNA6 (DFNA14, DFNA38) ± diabetes and optic atrophy[a] (Wolfram)	Tilted? (*tlt*)
	TFCP2L3	TFCP2L3 (transcription factor)	DFNA28	
	MTRNR1	Mitochondrial 12S rRNA	ND[a]	
	MTTS1	Mitochondrial tRNA$^{Ser(UCN)}$	ND[a]	

[a]Syndromic deafness.
[b]Subscript i denotes undefined locus number.
[c]Mutant obtained by gene knockout.
[d]Mutant obtained by targeted gene inactivation in the cochlear and vestibular sensory epithelia.
[e]Abbreviation: ND, not defined nomenclature.

involved in non-syndromic deafness in humans that have been identified to date is beyond the scope of this discussion. However, some of the genes underlying hereditary hearing loss that have been identified in recent years will be discussed to demonstrate the heterogeneity involved and the complexity of the hearing process. Genes will be discussed according to the nature of the defect they impose including: defects that impair the functioning of the delicate cochlear hair cells, that upset the ionic homeostasis of the inner ear and that impair the functioning of the tectorial membrane.

1.4.1 Hair cell defects.

Many abnormalities in hair cells that cause deafness result from cytoskeletal defects. For example, mutations in the genes encoding the unconventional (non-muscle) myosins (myosin VI, VIIA, and XV) all impair hair cell structure in some manner mainly by disrupting the organisation of the stereociliary hair bundle (for reviews see Petit et al, 2001; Redowicz, 2002; Ahmed et al, 2003). Unconventional myosin proteins are actin-based motor molecules that channel the energy of ATP hydrolysis to drive their motor along actin filaments. The structural integrity of the stereociliary hair bundle is paramount for mechanoelectrical transduction. Hence, mutations in the unconventional myosins underlie many forms of hereditary deafness. For example, mutations in myosin VI and XV underlie non-syndromic deafness loci DFNA22 and DFNB3, respectively. Whereas, mutations in myosin VIIA underlie both dominant (DFNA11) and recessive (DFNB2) forms of non-syndromic deafness. Mutations in myosin VIIA are also responsible for a syndromic form of hearing loss, Usher syndrome type 1B (for reviews see Petit et al, 2001; Redowicz, 2002; Ahmed et al, 2003). Usher syndrome (USH) is the most common cause of deafness that is associated with blindness. The shaker-1 mutant is a mouse model for Usher syndrome type 1B; the stereocilia are extremely disorganised and this leads to their eventual degeneration (for review see Ahmed et al, 2003). Similar to the unconventional myosins, harmonin is another important cytoskeletal component of hair cells and is located within the hair cell body and stereociliary hair bundle. Mutations in the gene encoding harmonin, a PDZ (Postsynaptic density protein, Disc-large, ZO-1) domain containing protein underlie Usher syndrome type 1C (Bitner-Glindzicz et al, 2000; Verpy et al, 2000) and DFNB18 (Jain et al, 1998). Recently, a number of investigations have reported that myosin VIIA together with several other USH type 1 proteins that have been identified to date including Cdh23 (USH1D), protocadherin 15 (USH1F) and SANS (USH1G) bind to PDZ domains within harmonin to form a dynamic functional 'protein network' in the stereocilia hair bundle that is essential for structural integrity (Boeda et al, 2002; Siemens et al, 2002; for reviews see El Amraoui and Petit, 2005; Reiners et al, 2006). This functional protein network is thought to be crucial for maturation of the hair cell stereocilia during development and may also be involved in the mechanoelectrical transduction process (for reviews see El Amraoui and Petit, 2005; Reiners et al, 2006.

1.4.2 Defects that impair the ionic homeostasis of the inner ear.

The ionic homeostasis of the inner ear is crucial for hair cell function. The endolymph surrounding the hair cell stereocilia is enriched in potassium ions and this is crucial to maintain the endocochlea potential essential for mechanoelectrical transduction (for review see Forge and Wright, 2002). Consequently, defects in molecules that play important roles in maintaining the ionic balance of the inner ear lead to deafness. Mutations in the voltage-gated potassium channel *KCNQ4* underlie DFNA2 and are thought to impair the potassium recycling in the inner ear (Coucke et al, 1999; Kubisch et al, 1999; Van Hauwe et al, 2000). Similarly, mutations in the *PDS* gene that encodes the protein pendrin are responsible for Pendred syndrome (the most common form of hereditary syndromic deafness) and DFNB4 (for review see Petit et al, 2001; Bitner-Glindzicz, 2002; Schrijver, 2004). Pendrin is an anion transporter and is proposed to function in endolymph homeostasis within the cochlear duct. It appears that mutations in *PDS* that abolish the transport of ions result in the syndromic form of deafness, whereas mild mutations that simply impair the transmission of ions manifest in the non-syndromic form (Scott et al, 2000).

Mutations in the genes encoding gap junction proteins that are expressed in the inner ear are also thought to impair the potassium recycling of the inner ear. Gap junction proteins are a major means of direct cell-cell communication and within the inner ear are found in abundance in the supporting cells population but are not associated with hair cells (for review see Forge and Wright. 2002). It is thought that the gap junctions enable the supporting cell population in the organ of Corti to channel potassium ions away from the endolymph surrounding the hair cell stereocilia. Three genes implicated in gap junction formation in the inner ear have been identified in hereditary deafness; these are the connexins genes: *GJB2* (DFNB1, DFNA3), *GJB6* (DFNB1, DFNA3) and *GJB3* (DFNAi) (for reviews see Rabionet et al, 2000; Petit et al, 2001; Gerido and White, 2004). Mutations in *GJB2* underlie around 50% of all cases of autosomal recessive, non-syndromic, congenital deafness in Caucasian populations. Mutations in two other connexin genes *GJB6* and *GJB3* do not show such a high prevalence in causing hearing loss (for reviews see Rabionet et al, 2000; Petit et al, 2001; Gerido and White, 2004). More than 50 different mutations have been identified in *GJB2*, the majority are recessive mutations that underlie DFNB1. However, some mutations present as dominant and underlie DFNA3. Certain mutations show a high prevalence depending on the ethnic background. The recessive mutation 35delG is very common in Caucasian populations particularly those of European-Mediterranean

descent, but has yet to be identified in Japanese, or African-American populations, and it appears rare in Asian populations. In contrast, the 167delT and 235delC mutations in *GJB2* are prevalent amongst Jewish and Japanese populations, respectively (for reviews see Rabionet et al, 2000; Petit et al, 2001; Gerido and White, 2004).

1.4.3 Tectorial membrane defects.

The tectorial membrane that overlies the sensory neuroepithelia of the organ of Corti is a gelatinous membrane composed of an extracellular matrix; it provides a support that facilitates deflection of the underlying hair cell stereocilia for mechanoelectrical transduction (for reviews see Forge and Wright, 2002; Hudspeth, 2005). Consequently, mutations in genes encoding components of this extracellular matrix are causative of hereditary hearing impairment. *TECTA* that encodes the α-tectorin protein is one such gene; it is responsible for both DFNA8/12 and DFNB21 (for reviews see Bitner-Glindzicz, 2002; Petit, 2005). Similarly mutations in *COL11A2* that encodes the α2 chain of the structural protein collagen X1, an important extracellular matrix component of the tectorial membrane are responsible for DFNA13 in addition to some syndromic forms of deafness (for reviews see Bitner-Glindzicz, 2002; Petit, 2005).

In summary, it is clear that many of the genes underlying syndromic and non-syndromic forms of hereditary hearing loss encode crucial components of the cochlea that are essential for maintaining structural integrity. Identification of the causative mutations responsible for monogenic forms of hereditary hearing loss have reaped tremendous advances in our understanding of the molecular basis of the hearing mechanism and indeed, hearing loss attributable to monogenic defects. However, in contrast to these monogenic congenital forms of syndromic and non-syndromic hearing loss very little is known about the genetic basis of late onset hearing loss, which is exhibited by a large proportion of the ageing population. Theoretically, any of the genes involved in congenital forms of hearing loss are candidates for susceptibility to late onset hearing loss; milder mutations may exist with phenotype of late onset. Understanding the pathophysiology and the genetics of late onset hearing loss is explored in detail in the next section (1.5).

1.5 Understanding the pathophysiology of late onset hearing loss.

1.5.1 Defining late onset hearing loss.

Historically, age-related hearing loss (ARHL) has been classified into four distinct categories based on audiometry and histological data: sensory, strial (metabolic), neural and conductive (Schuknecht, 1955; 1964; Schuknecht and Gacek, 1993). Under this framework laid down by Schuknecht, sensory ARHL is categorised physically by an elevation of hearing thresholds in the high frequencies and at the cellular level by a loss of hair cells in the basal turn of the cochlea. Strial ARHL is characterised by an elevation of hearing thresholds across all frequencies producing a flat audiogram and by degeneration of the stria vascularis. Neural ARHL is defined by a sloping audiogram with all frequencies reduced and a loss in the number of spiral ganglion neurons. Existence of conductive ARHL is hypothetical and is predicted to contain a descending sloping audiogram without any specific pathological associations. Hence, its existence remains controversial. In reality, there appears an overlap between these pathologies and convincing evidence for discrete forms of ARHL is lacking (Scholtz et al, 2001; for reviews see Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills, 2005). Indeed, the most common form of ARHL appears to be bilaterally symmetrical, most pronounced in the high frequencies and *sensorineural in origin.* Histopathologic investigations reveal that a progressive loss of hair cells initiating in the basal turn of the cochlea and proceeding along a basal to apical gradient is the predominant abnormality and this is usually coupled with a reduction in the number of spiral ganglion neurons. OHCs are more severely affected than IHCs; the OHC loss preceding IHC loss. Loss of spiral ganglion neurons where present appears secondary to hair cell loss (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller et al, 2004, Gates and Mills, 2005). It is this type of sensorineural hearing loss concomitant with ageing that is the focus of this project, which for clarity will be referred to as 'late onset hearing loss' and the general aim of this project is to investigate the genetic factors that underlie susceptibility with the intention to help elucidate why only some people retain good hearing into old age.

Late onset hearing loss is an extremely common and complex disease becoming increasingly common with advancing age. Similar to many other common diseases such as cardiovascular disease, diabetes, cancer and asthma, late onset hearing loss is regarded as a polygenic, multi-factorial disease, the aetiology of which is suspected to

involve a complex genetic-environmental interplay (for reviews see Jennings and Jones, 2001, Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills, 2005). Delineating the pathophysiology of late onset hearing loss is complex due to the multi-factorial nature of the disease. Genetic susceptibility is confounded by environmental factors including noise trauma and ototoxic drugs, and possibly nutritional health. Breaking all these factors down into discrete categories and understanding their individual contribution to development of late onset hearing loss is extremely difficult if not impossible. Drawing a clear distinction between hearing loss due to the 'ageing' process and chronic noise exposure is particularly problematic, as there appears much overlap between these two factors. Indeed, it is likely that ARHL and noise-induced hearing loss, NIHL (hearing loss caused by prolonged exposure to moderate level sound) are not separate pathologies but overlapping phenomena (for reviews see Ohlemiller et al, 2004; Gates and Mills, 2005). In agreement with this, it has been proposed that sensory ARHL as characterised by Schuknecht, 1955; 1964; Schuknecht and Gacek, 1993, in some cases at least, has more to do with the cumulative effect of acquired acoustic trauma than the intrinsic process of auditory ageing alone (for reviews see Ohlemiller et al, 2004; Gates and Mills, 2005). Consequently, for the purpose of this thesis the common, symmetric, sensorineural high frequency loss of hearing concomitant with advancing years was not defined by either of the commonly used terms: ARHL or presbyacusis. It is clear that ARHL is not just due to the inevitable process of ageing in the auditory system. Indeed, there is a large degree of inter-individual variability in onset of hearing loss and some people retain good hearing into old age (for reviews see Gates and Mills, 2005). In addition, there are conflicting reports in the literature regarding use of the term presbyacusis, which translated literally means 'elder hearing'. Presbyacusis has been described as 'the progressive sensorineural hearing loss associated with ageing' (Zheng et al, 1998; Alam et al, 2001) and in contrast, much more broadly, as 'the natural failure of hearing with advancing years caused by degenerative changes in the internal ear' (Jennings and Jones, 2001). Hence, for the purpose of this thesis and for clarity the term *late onset hearing loss* was adopted to describe the common, symmetric, sensorineural high frequency loss of hearing with advancing years to encompass the multifactorial nature of this disease. Late onset hearing loss was viewed as hearing loss manifesting from acquired environmental insults (for example, acoustic trauma and ototoxic agents) to the auditory system combined with the intrinsic process of auditory ageing. This concept of late

onset sensorineural hearing loss is in agreement with the current consensus as documented in the following recent reviews: Ohlemiller, 2004; Gates and Mills, 2005.

1.5.2 The genetic basis of late onset hearing loss.

At present genetic susceptibility to late onset hearing loss is poorly understood in humans. Monogenic traits including syndromic and nonsyndromic forms of congenital hearing loss often result from deleterious mutations within the coding regions of single genes that have a profound effect on protein function. In contrast, it is thought that moderate variations in gene expression or function at multiple loci underlie susceptibility to complex traits such as late onset hearing loss (for review see Fransen et al, 2003). A genetic basis underlying susceptibility to late onset hearing loss has long been suspected. This is illustrated from studies involving industrial workers who are exposed to the same noise over the duration of their career. For example, jute weavers are exposed to constant high-level noise (99-102 decibels) as an occupational hazard yet not all go on to develop late onset hearing loss induced by noise trauma; differences as great as 70dB have been reported between the least and most effected workers (Taylor et al, 1965). Hence, suggesting that genetic factors may be conferring increased susceptibility to chronic noise in the most affected workers. In addition, for some people at least, it is clear that susceptibility to late onset hearing loss is inherited. The heritability of late onset hearing loss is estimated to be in the region of 35 to 55% based on data scored from two studies: a Swedish population based study designed to examine the contributions of genetic and environmental factors on the development of ARHL in male twins (Karlsson et al, 1997), and an auditory study undertaken on members of the Framingham Heart Study to compare hearing threshold levels between genetically unrelated people (spouse pairs) and genetically related people (sibling pairs, parent-child pairs) (Gates et al, 1999).

In mice, steady progress is beginning to be made in unravelling the genetics of late onset hearing loss. Certain inbred strains of mice are prone to development of late onset sensorineural hearing loss (or ARHL as it is commonly referred to in mice) and considerable variation in onset and severity of the hearing loss is exhibited (Erway et al, 1993, Willott et al, 1998, Johnson et al, 2000). Three loci have been identified in mice that underlie susceptibility to ARHL: Ahl1 on mouse chromosome 10 (Johnson et al, 1997) and in more recent years, Ahl2 on mouse chromosome 5 (Johnson and Zheng, 2002) and Ahl3 on mouse chromosome 17 (Nemoto et al, 2004). The gene responsible

for Ahl1 has been identified as cadherin-23, Cdh23, (Bryda et al, 2001; Di Palma et al, 2001a; Noben-Trauth et al, 2003) and evidence, albeit controversial, suggests that this is an important structural component of the stereocilia tip link (Siemens et al, 2004; Sollner et al, 2004; Tsuprun et al, 2004; Lagziel et al, 2005; Michel et al, 2005). Currently, the genes underlying loci Ahl2 and Ahl3 have yet to be identified at the molecular level. However, once identified the genes responsible for Ahl2 and Ahl3 in mice would join Cdh23 (Ahl1) as good candidates for genetic susceptibility to late onset hearing loss in humans (Cdh23 as a candidate gene for late onset hearing loss in humans is discussed further in section 1.7.1).

At the onset of this project no common genetic variant had been identified as a risk factor for susceptibility to late onset hearing loss in humans. Although, a rare genetic variant in the Brn-3c gene had been found to be responsible for a monogenic form of adult onset hearing loss, DFNA15, in a large Israeli-Jewish family (Vahava et al, 1998). As such the Brn-3c gene represented an excellent candidate gene in which common sequence variants with a subtle effect on expression or function could underlie susceptibility to late onset hearing loss in the general population. The Brn-3c gene is the candidate gene chosen for the focus of this project and evidence that this gene is a good candidate gene for susceptibility to late onset hearing loss is reviewed in section 1.8.

1.5.3 The interaction between noise trauma and ageing in the auditory system.

It has been suggested that the effects of 'ageing' and 'chronic noise' on the cochlea contribute to the same pathology. That is, ARHL and NIHL (hearing loss caused by prolonged exposure to moderate level sound) are not separate pathologies but phenomena that overlap. In agreement with this proposal the impact of acoustic trauma and ageing can both be localised to the high frequency region of the cochlea and at the cellular level both NIHL and sensory ARHL are characterised by similar histological defects; a main feature of which is a loss of auditory OHCs (for reviews see Fransen et al, 2003; Ohlemiller et al, 2004). Furthermore, after a lifetime of chronic noise exposure ARHL and NIHL are reported to be practically indistinguishable audiometrically (for review see Fransen et al, 2003). However, the relationship between chronic noise exposure and ageing in the cochlea is equivocal and the mechanism by which these factors may induce the same pathology is far from clear. Several models have been proposed to explain the interaction between NIHL and ARHL (for review see

Rosenhall, 2003). The traditional, and perhaps simplest concept is that the effects of chronic noise exposure and ageing are 'additive'; each adds to the pathology (for review see Rosenhall, 2003). Indeed, it is thought that sensory ARHL as characterised by Schuknecht, 1955; 1964; Schuknecht and Gacek, 1993, has more to do with the cumulative effect of acquired acoustic trauma than the intrinsic process of auditory ageing alone (for reviews see Ohlemiller et al, 2004; Gates and Mills, 2005). An alternative concept is that chronic noise exposure and ageing interact to cause the pathology; that is, gene-environmental interactions underlie the pathology rather than just a purely additive effect between a gene(s) and the environment. Indeed, in mice an interaction between NIHL and ARHL is well documented and there is good evidence that noise exposure can accelerate onset of ARHL in genetically susceptible mice. For example, mice homozygous for the ARHL gene, Ahl1 ($Cdh23^{753A}$ / $Cdh23^{753A}$) develop ARHL but these mice exhibit an accelerated ARHL if they are exposed to acoustic trauma (Erway et al, 1996, Davis et al, 2001; Davis et al, 2003).

In practice, it is likely that both additive and non-additive effects as illustrated by Ahl1 homozygous mice are important in explaining the relationship between chronic noise exposure and ageing in the cochlea. Furthermore, the recent finding that *Cdh23* is also implicated in susceptibility to NIHL in mice (Holme and Steel, 2004) suggests a common molecular pathway(s) may be involved in the pathogenesis of NIHL and ARHL lending further support to the suggestion that ARHL and NIHL are not distinct pathologies.

1.5.4 Cellular and molecular mechanisms of late onset hearing loss.

It is clear that a complex genetic - environmental interplay underlies susceptibility to late onset hearing loss and that a major cause of this disease is loss of auditory hair cells (for reviews see Jennings and Jones, 2001, Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills, 2005). However, the aetiology of the disease that leads to the loss of these hair cells is as yet uncertain. Hair cell loss could be the result of pathological changes intrinsic to hair cells or it could be secondary to pathological changes extrinsic to hair cells; some of the possible molecular mechanisms intrinsic to hair cells will be discussed.

Many lines of evidence from studies with rodents suggest that apoptosis underlies the aetiology of hair cell loss concomitant with ageing and development of late onset hearing loss (Usami et al, 1997, Zheng et al, 1998, Alma et al, 2001). Zheng

et al, 1998 measured the distortion product of otoacoustic emission (DPOAE) in aged gerbils compared to young for an indication of outer hair cell function and in accordance with this analysis examined the incidence of apoptosis in the cochlea of these gerbils by the TUNEL (Terminal dUTP nick end labelling) assay. This analysis revealed TUNEL positive cells predominantly in OHCs compared to IHCs and the TUNEL positive cells correlated strongly with the aged gerbils that showed deterioration in cochlear function compared to the young control group and aged gerbils that showed no signs of deterioration in cochlea function (Zheng et al, 1998). Furthermore, immunohistochemical analysis on the cochlea in aged gerbils for specific members of the apoptosis cascade suggests that apoptotic cell death observed in the aged cochlea is due to suppression of bcl-2 protein expression and caspase-3 activation (Alam et al, 2001). Within the organ of Corti of aged gerbils, the number of bcl-2-immunostained cells was found to be significantly decreased along the basal, second, and apical turns and this was in accordance with a significant increase in the number of Caspase-3p20 immunostained cells. These results were not observed in the young gerbils, and were more marked in the aged gerbils that had severely deteriorated DPOAE responses suggestive of dysfunction at the OHC level (Alam et al, 2001). It is possible that the apoptotic loss of auditory hair cells is the end point of a molecular pathway(s) that underlies the pathophysiology of late onset hearing loss. Accordingly, genes involved in initiating or regulating apoptosis in hair cells are potential candidates for susceptibility to late onset hearing loss; common sequence variants in such genes could confer increased risk to this disease.

Ageing can be regarded an inherent process of cellular deterioration due to the progressive accumulation of metabolic and physiological changes (Partridge and Gems, 2002). Several theories have been proposed to explain the ageing process at the molecular level including: the dysdifferentiation theory of ageing, the telomerase theory of ageing and the 'mitochondrial clock theory' of ageing (also known as the membrane hypothesis of ageing). It is suggested that aspects of all these theories are applicable to the ageing process and indeed hearing loss associated with ageing (Seidman et al, 2002). Consequently, genetic susceptibility to late onset hearing loss could be precipitated by common sequence variants in genes involved in any of these molecular pathways. One theory to have received particular attention in recent years to explain hearing loss associated with ageing is the 'mitochondrial clock theory' of ageing (Seidman et al, 1997, 2000, 2002, 2004, Fischel-Ghodsian et al, 2004, Ohlemiller, 2004, Pickles, 2004). The 'mitochondrial clock theory' is based upon the generation of

reactive oxygen species, ROS (free radicals) leading to mitochondrial dysfunction. ROS are the extremely reactive by-products of oxidative metabolism produced by the mitochondria. One physiological feature of ageing is a reduced blood flow to many tissues including the cochlea and this results in increased ROS generation (Seidman et al, 1996). The ROS have the potential to damage mitochondrial DNA leading to specific deletions and ROS can lead to a reduction in the mitochondrial membrane potential with impaired mitochondrial function leading to yet more ROS production. Hence, the main tenet of this theory is the more damage that is accumulated the more the mitochondrial clock is accelerated to the point where the mitochondrion is rendered bioenergetically deficient. Therefore, long-term survival of auditory hair cells and maintenance of hearing may depend at least in part on the energy status of the mitochondria.

There is much evidence to support the involvement of ROS in the molecular pathogenesis of late onset hearing loss (McFadden et al, 1999a; McFadden et al 1999b; Ohlemiller et al, 1999; Staecker et al, 2001). One such line of evidence stems from superoxide dismutase-1, SOD1 (also known as copper/zinc superoxide dismutase, Cu/Zn SOD) knockout mice (McFadden et al, 1999a; McFadden et al 1999b). SOD1 is a cellular antioxidant enzyme that catalyzses the dismutation of the superoxide radical thus averting direct cellular damage via this radical (McCord and Fridovich, 1969). Mice knockout for the SOD1 gene exhibit increased susceptibility to ARHL; age-related changes in cochlea pathology including loss of auditory hair cells are more marked in SOD1 knockout mice compared to equivalent aged wild-type mice (McFadden et al, 1999a; McFadden et al 1999b) and this correlates with a significantly greater elevation in ABR thresholds at 8, 16 and 32kHz for SOD1 knockout mice compared to equivalent aged wild-type mice (McFadden et al 1999a). Interestingly, mice with targeted deletion of the SOD1 gene are also more susceptible to NIHL (Ohlemiller et al, 1999) suggesting that at least one molecular mechanism whereby ageing and acoustic trauma contribute to late onset hearing loss is mediated via superoxide radical generation. This is in agreement with the notion that ARHL and NIHL are not separate pathologies but phenomena united by a common molecular pathway(s).

A role for ROS in the molecular pathogenesis of NIHL is well documented (Ohinata et al, 2003; Yamashita et al, 2004; for review see Lynch and Kil, 2005). Interestingly, acoustic trauma to the cochlea is not the only exogenous stress that results in ROS generation; generation of ROS are also implicated in hair cell damage following use of ototoxic agents for example, cisplatin and aminoglycoside antibiotics (for reviews see Kopke et al, 1999; Campbell et al, 2003; Rybak and Whitworth, 2005). It is

suggested that oxidative stress inflicted on auditory hair cells via any of these environmental insults may underlie the apoptotic hair cell death that is observed in the cochlea when these exogenous factors are tested in animal models of hearing loss (for reviews see Duan et al, 2002; Lefebvre et al, 2002; Le Prell et al, 2003; Fischel-Ghodsian et al, 2004). It is tempting to speculate that these environmental insults on the cochlea act at least in part, by accelerating the demise of the mitochondrial clock. More recent evidence to illustrate the importance of functional mitochondria for maintenance of hearing into old age stems from transgenic mice that carry a proof reading-deficient version of the mitochondrial DNA polymerase γ (POLG); these mice accumulate mitochondrial mutations and exhibit an accelerated ageing including ARHL (Kujoth et al, 2005).

In summary, late onset hearing loss is an extremely common and complex disease attributable to a complex genetic-environmental interplay. The predominant type of late onset hearing loss in adults and exhibited by a large proportion of the ageing population is sensorineural in origin and a loss of cochlear hair cells, specifically the outer hair cells is one of the main histological defects. It is likely that apoptosis underlies cochlea hair cell loss and that survival is to some extent interrelated to the energy status of the mitochondria. In any case it seems clear that increased ROS generation whether through acquired auditory stress such as noise trauma and ototoxic agents or the inherent process of auditory ageing plays a crucial pathogenic role.

1.6 Methods for studying late onset hearing loss: a complex disease.

Similar to many other common diseases such as cardiovascular disease, diabetes, cancer and asthma, late onset hearing loss is an extremely common and complex trait. A complex trait can be defined as a 'disease where susceptibility is controlled by multiple genetic and environmental risk factors and where each of these risk factors has only a modest effect on susceptibility' (Cardon and Abecasis, 2003). Identifying the common genetic variants that predispose to complex disease such as late onset hearing loss is of immense importance; it will facilitate our understanding of the disease at the molecular level and provide a basis upon which therapeutic intervention to slow or halt disease progression can be developed. One model that has been proposed to explain genetic susceptibility to complex traits is the common disease: common variant (CD:CV) hypothesis (for reviews see Smith and Lusis, 2002; Wright et al, 2003). The principle of this model is that common genetic variants in the general population (with a rare allele frequency of or greater than 1%) underlie susceptibility to common complex disease. Hence, common alleles that predispose to the disease will be more prevalent amongst affected individuals than those who remain unaffected and these alleles have the potential to be identified in large-scale population based case-control association studies (Risch and Merikangas, 1996). However, delineating the genetic basis of late onset hearing loss as with all traits of a complex nature is difficult. Genetic susceptibility is confounded by environmental interplay and this is exacerbated by the polygenic nature of the disease; a plethora of factors are thought to interact to determine the final phenotype.

1.6.1 The nature of DNA sequence variation that underlies susceptibility to complex disease.

Genetic predisposition to complex disease such as late onset hearing loss is thought to result from subtle inter-individual variations in gene function and / or expression at multiple loci (for reviews see Smith and Lusis, 2002; Fransen et al, 2003; Wright et al, 2003). Consequently, the nature of DNA sequence variants that predispose to such complex traits are thought to have a mild effect on phenotype and these could manifest within any of the following regions of a candidate gene: the coding region, intron or regulatory region(s). For example, common sequence variants that occur within the

coding region of a candidate gene could manifest as synonymous single nucleotide substitutions that although may not change an amino acid in the protein product could affect alternative splicing of the gene. Indeed, synonymous single nucleotide substitutions have the potential to impair existing or create novel exonic splicing enhancer or silencer sites leading to an accumulation of both aberrantly and correctly spliced transcripts that have a mild affect on phenotype. In a similar manner, single nucleotide substitutions within intronic splicing enhancer or silencer sites can lead to a similar effect (for review see Nissim-Rafinia and Kerem, 2002). Alternatively, non-synonymous single nucleotide substitutions that result in amino acid substitutions could have a mild affect on phenotype if they occur outside of well-conserved protein domains that are crucial for functioning of the encoded protein.

Traditionally searches for sequence variants in disease genes have focussed on the coding region of candidate genes as opposed to the regulatory regions with the effect(s) of sequence variants in the promoter of a candidate gene often going uncharacterised and being excluded from analysis. However, sequence variants that manifest within the regulatory region of a candidate gene have the potential to result in subtle inter-individual varations in gene expression. For example, promoter variants can disrupt existing or create novel cis-acting elements and hence, modify the binding of transcription factors needed for gene regulation. Indeed, there is increasing evidence that sequence variants within gene regulatory regions underlie susceptibility to common, complex disease (Dawson et al, 1993; Zhang et al, 1999; Jormsjo et al, 2000; Niimi et al, 2002). In addition, the importance of promoter variants as a source of common inter-individual sequence variation is clearly evident from a recent large-scale systematic study into the extent of functional sequence variation that occurs within proximal promoter sequences (Hoogendoorn et al, 2003). Hoogendoorn et al examined the first 500bp of 5'-flanking promoter sequence in 170 human promoters and found that around 35% of the distal promoters screened contained common sequence variants, typically in the form of single nucleotide substitutions and of the promoter haplotypes examined approximately one third significantly altered transcriptional activity when tested in reporter gene assays (Hoogendoorn et al, 2003).

Furthermore, sequence variants that affect gene regulation are not just limited to the promoter of a gene they can manifest in the 3'- or 5'-untranslated region (UTR) of a gene. UTR sequence is transcribed to mRNA but is not translated to protein and is involved in the control of gene expression at the post-transcriptional level by affecting mRNA stability and translation (for reviews see Conne et al, 2000; Misquitta et al,

2001). Indeed, it is becoming clear that sequence variants that manifest within important cis-acting elements within the UTR of candidate genes have the potential to modifying the binding of proteins needed for gene regulation thus, leading to subtle inter-individual differences in gene expression that may be important in the pathophysiology of complex disease (Lambert et al, 2003; Hu et al, 2005; Wang et al, 2006). Therefore, when attempting to study complex disease and attempts are made to elucidate the genetic determinants responsible, it is important to consider the nature of DNA sequence variation that is likely to underlie susceptibility.

1.6.2 Genetic approaches to dissect the genetic factors that contribute to late onset hearing loss susceptibility: linkage and candidate gene case-control association analysis.

Two methods that have been widely used in an attempt to delineate the genetic basis of complex disease are: linkage analysis and the candidate gene case-control association approach (Lander and Schork, 1994, Douglas et al, 2001, Shifman et al, 2002, Botstein and Risch, 2003). Classically, linkage analysis examines the segregation of specific genetic markers with the disease phenotype in affected family pedigrees to map the disease locus in the genome; it makes no assumptions about the gene involved in disease susceptibility. If the map position can be narrowed down sufficiently then genes within the candidate region can be examined for the presence of mutations in affected subjects. Linkage analysis has been used very successfully in the identification of genes that have a major effect on disease phenotype and follow classic Mendelian inheritance patterns for example, in the identification of the many genes found to underlie monogenic forms of hereditary deafness (see Van Camp and Smith, Hereditary Hearing Loss Home Page, http://webhost.ua.ac.be/hhh/). A mutation that co-segregates only with affected individuals and is clearly deleterious to protein function is typically taken as evidence for involvement of the gene in the disease phenotype. However, in comparison linkage analysis has not been so successful in the detection of genes responsible for complex disease (for reviews see Risch, 2000; Altmuller et al, 2001). Complex diseases do not conform to classic Mendelian inheritance and consequently, linkage analysis is confounded by many factors including: genetic heterogeneity, incomplete penetrance, phenocopies, polygenic inheritance, and gene-environmental interactions (Lander and Schork, 1994). Furthermore, when the disease phenotype manifests later in life such as is the case with late onset hearing loss, the availability of

suitable large multi-generational families for linkage analysis that have surviving members of the elder generations is problematic and severely limits the progress that can be made using linkage analysis. However, using subjects originally recruited as part of the Framingham heart study and their offspring a large-scale genome wide linkage study to identify chromosomal loci that underlie susceptibility to hearing loss associated with ageing has been performed (DeStefano et al, 2003). Using DNA and audiometric data from 328 families spanning 2 or 3 generations six chromosomal loci were mapped, but linkage was not found at statistically significant levels. Interestingly, four of the loci mapped in this study overlapped with genes known to cause congenital deafness and of these three are involved in the Usher syndromes: Mysoin VIIa (Usher 1B), harmonin (Usher 1C) and the yet to be identified gene responsible for Usher 1A. It is possible that subtle mutations in these genes induce a moderate effect on development of late onset hearing loss. However, one anomaly with the study by DeStefano et al, is that it focussed on low and medium frequency hearing loss and not the high frequency hearing loss typical of sensorineural hearing loss commonly associated with ageing.

Currently, no loci that underlie susceptibility to the common sensorineural high frequency loss of hearing associated with ageing have been successfully mapped by linkage analysis. Indeed, evidence gained from other diseases suggests that using linkage analysis to identify the genetic determinants of complex traits of adult onset is only successful where a severe and clear definition of phenotype can be targeted for studies (for reviews see Lander and Schork, 1994; Risch, 2000; Botstein and Risch, 2003). This approach focuses on cases of the disease where there is the strongest genetic effect, which is therefore easier to detect. For example, identification by linkage analysis of the BRCA1 gene in breast cancer susceptibility was facilitated by only including subjects with an early age of onset in the study (Hall et al, 1990; Miki et al, 1994). In other more heterogeneous disorders where many factors both genetic and environmental may interact to produce the phenotype such as cardiovascular disease and schizophrenia, linkage analysis has proved much less successful (for reviews see Risch, 2000; Botstein and Risch, 2003).

Traditionally, case-control association studies are based on the pre-selection of a candidate gene using available pathophysiological knowledge about the disease under study and examining whether sequence variants in the candidate gene are associated with the complex trait. One advantage of using case-control association analysis rather than linkage analysis for the study of complex disease is that sample recruitment is often more practically achievable. Case-control association studies do not depend on the

recruitment of families that are affected with the disease but instead typically involve searching for genetic associations between a group of genetically unrelated people that have the disease compared to a control population of genetically unrelated people who do not have the disease. Hence, recruitment of for example, a large group of unrelated subjects with late onset hearing loss is more practically achievable than finding large families with multi generational affected elder surviving members. In this regard association analysis is not limited by late onset of disease phenotype to the same extent as linkage analysis. However, the main advantage of the case-control association study approach to detect genetic determinants of complex disease is that it provides far greater power than linkage analysis to detect genes of moderate effect (Risch and Merikangas, 1996). This is important given that the genetic determinants that contribute to late onset hearing loss are likely to be of moderate effect.

Although, a powerful approach to study complex disease the association study approach is not without its pitfalls. One confounding factor in association based analysis is that a sequence variant found to be at a significantly higher frequency in the 'case' population compared to the 'control' population in a single association study is not conclusive evidence for involvement of the sequence variant in the disease phenotype; false-positive associations can occur (type I error) (for reviews see Cardon and Bell, 2001; Hirschhorn et al, 2002; Donahue and Allen, 2005). Population stratification (the presence of multiple subgroups with different allele frequencies within a population) resulting from ethnic admixture (the presence of two or more different ethnic groups within a population) can lead to false-positive associations if the case and control populations are not ethnically matched (for reviews see Cardon and Bell, 2001; Hirschhorn et al, 2002; Donahue and Allen, 2005). Allele frequencies are known to vary amongst different ethnic groups; over-representation of a high-risk ethnic sub-group in the case population could lead to a significant difference in allele frequency between case and control populations being misinterpreted as evidence for involvement of the sequence variant in disease susceptibility. However, there are many ways to minimise this problem; most importantly measures should always be taken to obtain well-characterised case and control cohorts that are ethnically matched and population stratification, if present, can be corrected (Reich and Goldstein, 2001).

It must also be considered that a statistically significant difference in allele frequencies between case and control populations can arise not because the sequence variant is the casual allele but because it is a neutral variant in linkage disequilibrium with the casual allele at the same or another locus due to the tendency of closely linked

loci to co-segregate. However, by analysis of only functional sequence variants that have been established to have an affect on function or expression of the candidate gene under study one increases the likelihood that a positive association is due to identification of the causal sequence variant.

Historically, one of the main pitfalls of the candidate gene case-control association study approach to identify the genetic determinants that contribute to complex disease has been a lack of replication; many case-control association studies that have reported a positive association have failed to be independently replicated (for reviews see Cardon and Bell, 2001; Hirschhorn et al, 2002). Lack of replication can be due to many factors including but not limited to: type I error in initial study, type II error (false-negative association) in subsequent follow up study; often due to a lack of power, and problems in defining the phenotype of the disease under investigation between studies (for reviews see Cardon and Bell, 2001; Hirschhorn et al, 2002; Donahue and Allen, 2005). However, many of these anomalies can be reduced if not overcome.

Lack of replication in many small-scale case-control association studies to identify genetic determinants of complex disease which by nature are likely to have a weak effect on disease susceptibility, suffer from the 'winner's curse' phenomenon (Lohmueller et al, 2003). This phenomena arises because the first positive publication tends to over-estimate the extent of the genetic effect and subsequent follow up studies use too few samples (a hundred to a few hundred cases and controls or less) consequently they suffer from a lack of power and the association is not replicated. However, large sample sizes and/or a meta-analysis (pooling multiple independent studies together) can rectify this shortcoming (Altshuler et al, 2000; Keavney et al, 2000; Lohmueller et al, 2003). Indeed, as reviewed by Hirschhorn et al, 2002 the effect of the Pro12Al1 variant in the peroxisome proliferator-activated receptor-gamma gene (PPARγ) on type II diabetes risk was initially reported to be 3-fold. Subsequent studies failed to replicate this but follow-up studies using large populations revealed that the risk was much smaller, 1.25-fold. Similarly, the need for association studies to use large sample sizes to achieve consistency with replication has been shown by analysis of sequence variants in the angiotension-converting enzyme (ACE) gene and risk to cardiovascular disease (Keavney et al, 2000). Furthermore, as emphasised by Rebbeck et al, 2004 the lack of replication in many association studies may be because many of these use sequence variants that are uncharacterised. A variant that is linked to a functional variant in one population may not be linked in another and will therefore not

produce a significant association. By only using functional sequence variants that are known to have an effect on function or expression of the candidate gene under study, the likelihood of identifying the causal variant is increased and therefore so is the likelihood of replication in different samples.

Importantly, despite earlier problems the case-control association study approach focussing on the pre-selection of a candidate gene has already been used very successfully to identify genetic risk factors in complex traits including Alzheimer's disease [apolipoprotein E, APOE] (Corder et al, 1993), type II diabetes [peroxisome proliferator-activated receptor-gamma, PPARγ] (Deeb et al, 1998; Altshuler et al, 2000) and deep vein thrombosis [factor V] (Bertina et al, 1994; Dahlback, 1997). Undeniably, it has been widely applied in delineation of the genetic risk factors for cardiovascular disease (for reviews see Daley and Cargill, 2001; Green, 2001; Herrmann and Paul, 2002; Humphries et al, 2004). For example, a 4G/5G insertion / deletion sequence variant at -675 in the plasminogen activator inhibitor type 1 (PAI-1) promoter appears to modify binding of a transcriptional repressor and is thought to be important in regulation of the PAI-1 gene (Dawson et al, 1993); this sequence variant has consistently been associated with susceptibility to thrombotic disease (for reviews see Nordt et al, 2001; Dellas and Loskutoff, 2005).

At present the candidate gene case-control association study approach appears one of the most suitable methods with which to delineate the genetic basis of late onset hearing loss. This is especially so given that it provides far greater power than linkage analysis to detect genes of modest effect (Risch and Merikangas, 1996). For example, as highlighted by Hirschhorn and Daly, 2005 the common Pro12Ala variant in PPARγ was found to affect the risk of type II diabetes by case-control association analysis but this would never have been achieved by linkage analysis; at least one million affected sib-pairs would be needed to achieve statistical significance using linkage. However, this does not mean that linkage analysis should not be considered for analysis of complex traits. It is probably more appropriate to view linkage and candidate gene case-control association analysis as somewhat complementary approaches to study genetic susceptibility to complex disease. Large-scale genome wide linkage analysis using affected sib-pairs has provided candidate chromosomal loci from which case-control association analysis performed within the candidate region has been used to identify the genetic variant associated with the complex trait; using this approach genetic variants in the ADAM33 gene have been associated with susceptibility to asthma (Van Eerdewegh et al, 2002).

1.6.2.1 The International Haplotype Map (HapMap) project.

Recently, the International Haplotype Map (HapMap) project has been formed with the aim to characterise the common patterns of DNA sequence variation in the human genome (International HapMap Consortium, 2003). This will in effect catalogue the extent of linkage disequilbrium throughout the entire genome as a series of haplotype blocks. A haplotype block is fundamentally a chromosomal region of high linkage disequilibrum and limited haplotype diversity. A limited number of common haplotypes account for all common haplotypes within the block and these can potentially be identified by genotyping a selected number of SNPs, so named 'tag-SNPs' (for reviews see Johnson et al, 2001; Cardon and Abecasis, 2003). Large-scale genome wide association studies using tag-SNPs are anticipated upon completion of the HapMap and this approach is predicted to be of immense importance in facilitating our understanding of the genetic factors that contribute to complex disease (for review see Farrall and Morris, 2005; Hirschhorn and Daly, 2005). Indeed, it will certainly be useful in identifying novel genetic risk factors in genes that have yet to be identified or well-characterised and hence, are not currently implicated in late onset hearing loss susceptibility. However, the outcome of such large-scale genome wide association studies is some time away and many important decisions regarding the design of the study, the analysis and the interpretation of the large sets of data that will no doubt accumulate still need to be made. For example, the exact number of tag-SNPs that will need to be genotyped in order to provide information about the majority of common variants in the human genome is unclear at present; the most recent estimates suggest a few hundred thousand for subjects of European descent with much larger numbers likely for subjects of African descent (for review see Hirschhorn and Daly, 2005). Ultimately, the number will be subject to many variables; the efficiency of tag-SNPs in regions of low linkage disequilibrium is just one such consideration (for review see Hirschhorn and Daly, 2005; Wang et al, 2005). Furthermore, genotyping all these tag-SNPs in each case and control individual is a most challenging and costly undertaking. Especially given that sample sizes need to be large (several thousand cases and controls) to achieve high statistical power not only because the genetic determinants that contribute to complex traits are likely to induce moderate effects but because one must account for the multiple-hypothesis testing when using a genome-wide association approach (for reviews see Farrall and Morris, 2005; Hirschhorn and Daly, 2005; Lawrence et al, 2005; Wang et al, 2005). Indeed, Hirschhorn and Daly propose that pilot experiments are needed to test the merits of the genome-wide association study

approach before numerous costly studies are undertaken. Finally, even when positive associations to for example, haplotype blocks are found using this genome-wide association study approach the causal variant implicated in disease susceptibility will still need to be identified with *in-vitro* and/or *in-vivo* functional analysis. It is only by gaining knowledge of the functional effect(s) of a genetic variant that their causative role in disease pathogenesis can be fully assessed and a greater understanding of the cellular and molecular mechanisms that underlie disease susceptibility can be established. Consequently, the genome wide association study approach should be considered a complementary method to the traditional candidate gene case-control association study approach. The genome wide association study approach provides a long-term, future method to identify novel late onset hearing loss susceptibility genes whereas the current candidate gene approach provides a quicker route to determine the effect of known candidate genes on susceptibility to late onset hearing loss.

1.6.3 Quantitative trait analysis to dissect the genetic factors that contribute to late onset hearing loss susceptibility.

The traditional case-control association study approach to identify the genetic determinants that contribute to complex disease involves stringent classification of the phenotype under study into discrete forms. One is either classified as having the disease under study or not. A drawback of this approach to detect the genetic variants that underlie late onset hearing loss susceptibility is that it dichotomises what can be viewed as a quantitative trait; deterioration of hearing with advancing age is extremely common and it is whether this deterioration reaches levels which affect the individual's daily life that determines whether they become classified as a patient. Dichotomising a quantitative trait results in a reduction of statistical power (Page and Amos, 1999).

Fransen et al, 2003 have proposed that it would be more useful to define a continuous parameter for ARHL that represents the degree of hearing ability in a given subject compared to a median value and to use this continuous parameter to enable analysis of ARHL as a quantitative trait. Hence, rather than dichotomising subjects as either having late onset hearing loss or not and trying to find significant differences in allele or genotype frequencies between these two discrete populations one would seek to find associations between sequence variants in a candidate gene and severity of late onset hearing loss. Recently, in order to analyse ARHL as a quantitative trait Fransen et

al, 2004 have devised a novel method based upon defining a quantitative trait value for ARHL.

In defining a quantitative trait value for ARHL Fransen et al, 2004 used a Z-score approach to describe the extent by which an individual is affected by ARHL. Fransen et al, used a Z-score to represent how frequency specific hearing thresholds for a given individual differ from median values for a particular age and gender (Fransen et al, 2004). This approach is appealing, as it will enable genetic analysis of ARHL as a quantitative trait with greater statistical power to detect genes of modest effect as opposed to a binary trait concurrent with a loss of statistical power due to discrete 'case' versus 'control' classification of what is effectively a more continuous than discontinuous parameter. To illustrate this approach Fransen et al, genotyped a random sample of 101 subjects for a common sequence variant in the cochlin (COCH) gene, the gene, which underlies DFNA9 (Robertson et al, 1998) and converted audiometric data obtained for these subjects into a series of Z-scores. No association between extent of hearing loss (Z-score) and COCH genotype was found in this pilot study, but the pilot study represented a random sample of subjects and was based on a relatively small sample size. Indeed, it is clear from the study by Fransen et al, 2004 that with larger sample sizes (Fransen et al, 2004 propose 500-1000 samples) the Z-score approach to analyse late onset hearing loss as a quantitative trait could be a promising method with which to delineate the genetic factors that underlie susceptibility.

In summary, the case-control association study approach offers some clear advantages to delineate the genetic basis of late onset hearing loss compared to linkage analysis; a main advantage of which is that it provides far greater power than linkage analysis to detect genes of modest effect (Risch and Merikangas, 1996). The large-scale genome wide association studies that will emerge in the future upon completion of the HapMap will certainly facilitate identification of the genetic determinants responsible for late onset hearing loss susceptibility. However, these results are some time away and at present carefully designed case-control association studies using well-characterised patient cohorts and focussing on functional sequence variants in the genes that are already known to be good candidates for late onset hearing loss susceptibility offer an instant and promising approach. Furthermore, recent developments by Fransen et al providing a means by which to analyse late onset hearing loss as a quantitative trait suggest that this novel method to analyse candidate genes for late onset hearing loss susceptibility will be a future, powerful approach.

1.6.4 The mouse as a model organism for late onset hearing loss.

There is strong similarity between the structure and functioning of the auditory system in humans and mice and the mouse, as a model organism, has proved very useful for studies of hearing and hearing loss in humans. Specifically, 'comparative genomics' between human and mouse genomes has lead to the elucidation of many genes involved in congenital forms of non-syndromic deafness in humans (for reviews see Probst and Camper, 1999; Kiernan and Steel, 2000; Ahituv and Avraham, 2002; Avraham, 2003). Consequently, use of mice as a model organism to facilitate our understanding of late onset hearing loss pathogenesis in humans is an appealing approach, although not without its limitations.

The problem with dissecting the pathophysiology of late onset hearing loss in humans is that the aetiology of this disease is extremely complex; a genetic-environmental interaction is thought to underlie susceptibility and deciphering the effects of ageing on the auditory system in humans from confounding environmental factors such as noise trauma is difficult. This is further compounded given that it seems ARHL and NIHL are not separate pathologies but over lapping phenomena (for reviews see Ohlemiller et al, 2004; Gates and Mills, 2005). Hence, one advantage of using the mouse as a model organism for late onset hearing loss is that the effects of ageing and noise exposure on the auditory system can be studied separately. Mice are easy to manipulate and experiments can be performed in a quiet or noise-controlled environments as appropriate. Furthermore, mice have a short life-span (approximately three years at most) and the effect of the ageing process on the auditory system can be observed within a small time frame. In addition, use of inbred strains reduces inter-laboratory experimental variation when experimental data is compared across different laboratories. It also minimises intra-experimental variation when experiments are repeated as the complication of a variable genetic background is removed (for review see Avraham, 2003). On the other hand the limited life-span of the mouse coupled with the nature of the environment in which they tend to be raised (quiet, free from exogenous stress) can also be regarded a clear disadvantage. Late onset hearing loss in humans can take many years to manifest (for example, with onset at 50, 60, 70 years and above) and the extent of exogenous stress encountered, especially noise trauma is an important factor in the pathogenesis. Furthermore, the genetic determinants of late onset hearing loss disease susceptibility are likely to have subtle effects; such subtle effects are extremely unlikely to be identified in 2 to 3 year old mice that have been raised in a quiet, controlled environment.

Despite clear inter-species differences between mice and humans, identification of the genes responsible for and / or that promote ARHL and NIHL in mice can provide clues as to which genes may be involved in late onset hearing loss susceptibility in humans. It has been known for some time that depending on the genetic background inbred strains of mice are prone to development of late onset sensorineural hearing loss, often termed 'Ahl' (Erway et al, 1993; Willott et al, 1998; Johnson et al, 2000). One such inbred strain that is extremely well characterised is C57BL/6J; late onset sensorineural hearing loss is inherited in these mice as a recessive trait (Erway et al, 1993; Johnson et al, 1997; Parham, 1997; Spongr et al, 1997). C57BL/6J mice develop high-frequency sensorineural hearing loss that manifests around 3-6 months of age and progresses to a severe hearing loss by 1 year of age. Pathologically, histological analysis shows that OHCs of the cochlea are more susceptible than IHCs and degeneration progresses along a basal to apical gradient. At 6 months of age degeneration of spiral ganglion neurons is also evident (Spongr et al, 1997). This phenotype is very similar to that of late onset sensorineural hearing loss that has been described in humans (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller et al, 2004, Gates and Mills, 2005). Furthermore, C57BL/6J mice are also susceptible to NIHL (Erway et al, 1996; Davis et al, 2001); all these features make the C57BL/6J mouse strain a promising model for at least some of the genetic factors that may contribute to late onset hearing loss susceptibility in humans. Indeed, genetic mapping in C57BL/6J mice has identified the locus responsible for late onset sensorineural hearing loss, Ahl1 (Johnson et al, 1997) and this gene has recently been identified as Cadherin-23, Cdh23 (Bryda et al, 2001; Di Palma et al, 2001a; Noben-Trauth et al, 2003). Cdh23 appears to be an important component of stereocilia tip-links (Siemens et al, 2004; Sollner et al, 2004; Tsuprun et al, 2004; Lagziel et al, 2005; Michel et al, 2005) raising the possibility that this gene or other genes involved in stereocilia integrity could underlie late onset hearing loss susceptibility in humans. Cdh23 as a candidate gene for late onset hearing loss in humans and additional loci for late onset sensorineural hearing loss that have been mapped in mice are discussed further in section 1.7.1.

Importantly, understanding how genes involved in ARHL and NIHL function in mice and the mechanism by which they cause or promote ARHL and NIHL in these animals (for example, by additive effects or by epistatic interactions that may or may not be dependent on the environment) can give insight into the underlying mechanisms that are responsible for late onset hearing loss in humans. For example, in CBy-dfw ^{2J}/+

mice it is clear that an epistatic interaction between Ahl1 (Cdh23) and PMCA2 (a calcium pump localised to stereocilia hair bundles) (Yamoah et al, 1998; Dumont et al, 2001) can accelerate onset of the ARHL phenotype in these mice (for review see Davis et al, 2003; see also section 1.7.2 for more detail). Therefore, it is possible a similar interaction between Cdh23 and PMCA2 or at least, genes involved in stereocilia integrity and maintenance of ionic homeostatsis in the cochlea may be applicable to development of late onset hearing loss in humans (this is discussed in section 1.7.2).

The ease with which mice can be manipulated genetically is a powerful tool for aiding our understanding of the cellular and molecular pathways that underlie late onset hearing loss susceptibility. Targeted gene deletion to create 'knockout' mice with null copies of the gene of interest means the function of any gene and hence, its potential role in late onset hearing loss susceptibility can effectively be examined. This technique does however, have its limitations. Functional redundancy within gene families can mask the effect of the knockout gene under study and knockout of some genes is limited by embryonic lethality. Nevertheless, it is clear that by loss of function studies some genes either increase susceptibility to NIHL or ARHL in mice. For example, targeted deletion of the PMCA2 gene increases susceptibility to NIHL in PMCA2 +/- heterozygotes (Kozel et al, 2002) whereas, targeted deletion of the Barhl1 gene (a homeobox transcription factor) in Barhl-/- mice leads to profound deafness due to degeneration of inner and outer hair cells in the organ of Corti that is progressive with age (Li et al, 2002). Moreover, it is clear that in mice loss of function of some genes increases susceptibility to both NIHL and ARHL for example, the cellular antioxidant enzyme SOD1 (McFadden et al, 1999a; McFadden et al, 1999b; Ohlemiller et al, 1999; McFadden et al, 2001) raising the notion that ARHL and NIHL may not be separate pathologies but overlapping phenomena (all these genes are explored further in section 1.7 as candidates for late onset hearing loss in humans).

In summary, it is clear that the mouse is a valuable model organism for study of late onset hearing loss in humans not least because of some practical benefits but because the nature of genes involved in ARHL or NIHL in mice provide candidates for the study of late onset hearing loss in humans. In addition, by elucidating the mechanisms by which genes involved in ARHL and NIHL in mice function clues can be gained into the underlying mechanisms that are responsible for late onset hearing loss in humans. However, it is important to remember that mice and humans are entirely different species and at present none of the genes involved in ARHL or NIHL in mice have been convincingly identified in late onset hearing loss susceptibility in humans.

1.7 Candidate genes for late onset hearing loss.

Late onset hearing loss is an extremely complex pathology and becomes increasingly common with advancing age. Susceptibility to late onset hearing loss, as with many traits of a complex nature is thought to be due to subtle variations in gene expression or function at multiple loci coupled with environmental interplay (for reviews see Jennings and Jones, 2001, Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills, 2005). In principle, genetic variation in a candidate gene could act to amplify or to reduce risk of developing late onset hearing loss and the degree of risk in any one individual is likely to be subject to the nature of epistatic interactions and exogenous stress encountered. This forms the basis for late onset hearing loss as a multifactorial disease. In theory, any of the genes expressed within the inner ear and crucial to auditory function is a candidate for susceptibility to late onset hearing loss; milder phenotypes may exist for some of the many genes involved in congenital forms of non-syndromic sensorineural hearing loss. However, the predominant cellular defect associated with late onset hearing loss is a progressive loss of hair cells from the cochlea (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller et al, 2004, Gates and Mills, 2005). Accordingly, any of the various genes involved in maintaining hair cell structure and function (for reviews see Steel and Kros, 2001, Petit et al, 2001) are particularly good candidates for susceptibility to this disease.

At the onset of this project very few genes were implicated as strong candidate genes for susceptibility to late onset hearing loss in humans. One of the most promising candidates was the POU domain transcription factor Brn-3c. Several lines of evidence had accumulated to suggest that Brn-3c functions in maturation and maintenance of auditory hair cells during development and ultimately as a life-long pro-survival factor for these sensory hair cells (Erkman et al, 1996, Xiang et al, 1997, Xiang et al, 1998 and Vahava et al, 1998). Given that loss of auditory hair cells appears to be the main cause of late onset hearing loss, these features make Brn-3c a very strong candidate gene for susceptibility to this disease (evidence to implicate Brn-3c as a good candidate gene for late onset hearing loss is reviewed in section 1.8). During the course of this project several additional promising candidate genes for late onset hearing loss have emerged and case-control association analysis has been performed on a few of these (Fortunato et al, 2004; Unal et al, 2005). It is not possible to discuss exhaustively all possible candidate genes for late onset hearing loss but some of the more promising candidate

genes that have emerged in recent years and join Brn-3c as a good candidate gene for susceptibility to this disease will be discussed below.

1.7.1 Candidate genes for late onset hearing loss that have emerged from mapping experiments in mice: with emphasis on Cadherin-23.

Probably some of the strongest candidate genes for susceptibility to late onset hearing loss in humans have emerged from mapping experiments in mice. Certain inbred strains of mice are particularly susceptible to late onset sensorineural hearing loss (which is often termed Ahl) (Erway et al, 1993; Willott et al, 1998; Johnson et al, 2000) and currently three loci have been mapped termed: Ahl1, Ahl2 and Ahl3 on mouse chromosome 10 (Johnson et al, 1997), 5 (Johnson and Zheng, 2002) and 17 (Nemoto et al, 2004), respectively. The gene responsible for Ahl1 is the first gene causing non-syndromic hearing loss of late-onset that has been identified and mapped in the mouse (Johnson et al, 1997) and has long been known to be inherited as a recessive trait in the C57BL/6J inbred mouse strain (Erway et al, 1993). In addition, Ahl1 is implicated in late onset hearing loss exhibited by at least 10 different strains of mice (Johnson et al, 2000). The human orthologoue of Ahl1 and the recently identified loci, Ahl2 and Ahl3, are all good candidates for susceptibility to late onset hearing loss in humans. Although at present, the genes responsible for Ahl2 and Ahl3 have yet to be identified at the molecular level. However, several candidate genes have been proposed for Ahl1 (Johnson et al, 1997, 2000) and recent evidence indicates that Cadherin 23 (Cdh23) underlies Ahl1 (Bryda et al, 2001, Di Palma et al, 2001a; Noben-Trauth et al, 2003). Cadherins are transmembrane proteins important in cell-cell interactions and within the cochlea Cdh23 is crucial for organization of the stereocilia array (Di Palma et al, 2001a). Recent evidence suggests that Cdh23 is a main component of the tip-link joining adjacent stereocilia (Siemens et al, 2004; Sollner et al, 2004; Tsuprun et al, 2004). Although, this evidence is somewhat controversial and has been contradicted by others (Lagziel et al, 2005; Michel et al, 2005) who report that Cdh23 is absent from the apex of stereocilia in adult mice.

Evidence that Cdh23 underlies Ahl1 stems from the identification of mutations in Cdh23 as underlying the waltzer mutation in the mouse (Di Palma et al, 2001a); waltzer ($Cdh23^{v}$) maps to the same region of mouse chromosome 10 as Ahl1 (Bryda et al, 1997, Bryda et al, 2001) suggesting they are allelic. In an attempt to elucidate mutations in Cdh23 that underlie Ahl1 82 single nucleotide substitutions have been

identified based on sequence comparison between the coding and 5'- and 3'-untranslated regions in C57BL/6J and CAST/Ei strains of mice (CAST/Ei retain good hearing into old age) (Di Palma et al, 2001b). Theoretically, any of these could be responsible for Ahl1 in C57BL/6J mice. Di Palma et al, 2001b ruled out 10 single nucleotide substitutions that led to amino acid substitutions in the coding region of Cdh23 as the cause of Ahl as these also appeared in the mouse strain CBA/CaJ that retain hearing throughout their life. Similarly, single nucleotide substitutions within the 5'- and 3'-untranslated regions of Cdh23 and an isoform of Cdh23 with a spliced exon 68 that is predominantly expressed in the organ of Corti, have been excluded as the basis of the Ahl1 phenotype in C57BL/6J mice (Di Palma et al, 2001b). Di Palma et al, 2001b suggested that it may be one of the 61 apparently "silent" single nucleotide substitutions that do not alter the amino acid sequence that predispose to Ahl1 in C57BL/6J mice. It was predicted that these single nucleotide substitutions could affect exon splicing by disrupting exonic splicing enhancer or silencer motifs. This is certainly possible, it is known that 10-15% of disease causing mutations in human genes affects pre-mRNA splicing (Nissim-Rafinia et al, 2002). Recent evidence is in agreement with this suggestion; a synonymous single nucleotide substitution in exon 7 of Cdh23, $Cdh23^{753A}$, that causes in-frame skipping of exon 7 has been found to exhibit significant association with Ahl1 (Noben-Trauth et al, 2003). Twenty-seven of thirty-one inbred mouse strains characterised with ARHL were found to carry the variant allele, $Cdh23^{753A}$, and of 25 inbred mouse strains that do not develop ARHL, 22 carried the wild-type allele, $Cdh23^{753G}$ (Noben-Trauth et al, 2003).

In humans different allelic forms of Cdh23 underlie both the recessive syndromic-hearing loss trait Usher syndrome type 1D (Bolz et al, 2001) and non-syndromic autosomal recessive deafness DFNB12 (Bork et al, 2001). Hence, it is certainly feasible that mild allelic variants of Cdh23 may contribute to late onset hearing loss in the general population. Recent evidence certainly suggests that this may be so (Yang et al, 2006). A small-scale case-control association study on 194 noise-exposed Chinese workers comprising 93 workers with NIHL and 101 workers with normal hearing as determined by audiometric analysis revealed an association between three sequence variants in the Cdh23 gene and NIHL (Yang et al, 2006). It would be interesting to see if these associations are also detectable in noise-exposed Caucasians or indeed Caucasian subjects classified with late onset hearing loss of sensorineural origin. In any case it is clear that the study by Yang et al requires confirmation in a larger cohort for validation.

1.7.2 The genetic interaction between Cdh23 and PMCA2: relevance to late onset hearing loss pathology.

In 1998, Street et al showed that mutations in PMCA2 are responsible for the congenital deafness that underlies the deaf waddler (dfw) phenotype in the mouse; these mice are deaf, exhibit a wobbly gait when walking and display head bobbing (Street et al, 1998). In the cochlea PMCA2 is localised to the stereocilliary hair bundle in particular, the stereocilia of OHCs (Dumont et al, 2001). PMCA2 functions to exclude calcium from the stereocilia (Yamoah et al, 1998) and appears crucial in maintaining the calcium rich endolymph surrounding the stereocilia necessary for proper functioning of the hair cells (Wood et al, 2004). Two mutant alleles of dfw exist; both arose spontaneously in mice of different strains. An amino acid substitution in a conserved residue underlies the dfw allele (that originated in mice of the C3H/HeJ strain), whereas a 2bp deletion is predicted to cause a truncated protein in the second allelic form, dfw^{2J} (that originated in a substrain of BALB/cByJ mice termed, CBy-dfw^{2J}) (Street et al, 1995, Noben-Trauth et al, 1997, Street et al, 1998). Both of these mutations present as recessive. However, depending on the genetic background of the mice, a modifier gene termed "modifier of deaf waddler" (mdfw) can precipitate hearing loss in dfw^{2J} heterozygotes via an epistatic intereaction (Noben-Trauth et al, 1997). The auditory brainstem response (ABR) of CBy dfw^{2J}/+ mice are abnormal, elevated thresholds manifest that show a progressive dependence on age; hearing loss initiates four weeks after birth progressing to complete deafness by 12 weeks of age (one would expect CBy dfw^{2J}/+ mice to have normal hearing consistent with the recessive nature of the mutation).

An elegant series of experiments have shown that the early onset progressive hearing loss in CBy dfw^{2J}/+ mice is due to homozygosity with respect to Ahl1; Ahl1 (Cdh23) is mdfw. Indeed, evidence indicates that mdfw is allelic to both Ahl1 and $Cdh23^{v}$ that underlies the waltzer mutation in the mouse (Byrda et al, 2001; Zheng and Johnson, 2001). Hence, suggesting that the Ahl1, mdfw, and waltzer loci all represent the same gene, Cdh23. In agreement with this, the synonymous single nucleotide substsitution identified in exon 7 of Cdh23, 753A, that correlates with Ahl1 also segregates with mdfw in CBy-dfw^{2J}/+ mice in addition to all mouse strains that show allelism between the Ah11 and mdfw loci (Noben-Trauth et al, 2003). Hence, mice homozygous for the Ahl1 gene ($Cdh23^{753A}$/ $Cdh23^{753A}$) exhibit ARHL but an earlier ARHL is exhibited in these mice if they also possess a mutant copy of PMCA2 termed, dfw^{2J}. Calcium binding is known to be essential for cadherin function. Thus, it is

conceivable that impaired calcium homeostasis resulting from mutations in PMCA2 could accelerate onset of the $Cdh23^{753A}$/ $Cdh23^{753A}$ phenotype.

It is possible that an epistatic relationship between Cdh23 and PMCA2 as observed in the mouse is applicable to development of late onset hearing loss in humans. Schultz et al, 2005 have recently reported a genetic interaction between *CDH23* and *PMCA2* underlies the hearing loss in a Caucasian family that presents with autosomal recessive, non-syndromic sensorineural hearing loss manifesting in the first decade of life. In this family all affected members exhibit severe to profound high-frequency hearing loss that has been attributed to a homozygous mutation in *CDH23* (F1888S) but for some members the hearing loss phenotype is much more severe affecting all frequencies and this is proposed to be due to a mutation in *PMCA2* termed, V586M. Hence, it is certainly likely that mild variants in *CDH23* may exist that confer susceptibility to late onset hearing loss in the general population and it is possible individuals homozygous for a mutant allele of *CDH23* may be particularly susceptible to late onset hearing loss if they also possess a mutant copy of *PMCA2*.

In addition, both Cdh23 and PMCA2 are known to confer susceptibility to NIHL in mice (Erway et al, 1996, Davis et al, 2001; Kozel et al, 2002; Holme and Steel, 2004. It has been known for some time that mice that have a genetic predisposition to ARHL due to homozygosity at Ahl1 are also more susceptible to NIHL (Erway et al, 1996, Davis et al, 2001). Similarly, both mice heterozygous for a presumed null mutation in Cdh23 (Holme and Steel, 2004) and a null allele of PMCA2 (Kozel et al, 2002) are susceptible to NIHL. This raises the possibility that individuals carrying mild allelic variants in *CDH23* or *PMCA2* may be more vulnerable to the effects of acoustic trauma. Indeed, following acoustic over-stimulation the intracellular calcium concentration of OHC is known to be elevated (Fridberger et al, 1998) and it is proposed that NIHL exhibited by PMCA2+/- mice is due to a reduced ability to counteract any rises in cytoplasmic calcium which in turn is deleterious to the functioning of the OHCs (Kozel et al, 2002). Furthermore, it is known that stereocilia can be flattened by exposure to excessive noise (Wang et al, 2002). Cdh23 is paramount for organisation of the stereocilia array (Di Palma et al, 2001a) and is thought, albeit somewhat controversially, to be an important structural component of stereocilia tip-links (Siemens et al, 2004; Sollner et al, 2004; Tsuprun et al, 2004; Lagziel et al, 2005; Michel et al, 2005). Hence, it is certainly conceivable how subtle variant(s) in *PMCA2* and / or *CDH23* in the general population that impair the functioning or expression of these genes could predispose to late onset hearing loss precipitated by acoustic trauma.

In summary, it is clear that either individually or in combination that both *CDH23* and *PMCA2* represent excellent candidate genes for susceptibility to late onset hearing loss in humans. Furthermore, the evidence gained from mice suggests that if allelic variants in *CDH23* and *PMCA2* are present in combination and an epistatic interaction between these two genes predisposes to late onset hearing loss in humans, onset of the hearing loss phenotype could be greatly accelerated by exposure to acoustic trauma.

1.7.3 Barhl1 a hair cell survival factor is a good candidate gene for late onset hearing loss.

Loss of inner ear sensory hair cells appears to be the main cause of late onset hearing loss (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller et al, 2004, Gates and Mills, 2005). Accordingly, any genes involved in hair cell survival are very good candidates for susceptibility to late onset hearing loss. It has become apparent that Barhl1, which encodes a transcription factor, is required for long-term survival of sensory hair cells (Li et al, 2002). Barhl1 belongs to the Bar subset of homeobox genes and is closely related to the *Drosophila* BarH1 and BarH2 genes that are important in development of the eye and external sensory organs (Bulfone et al, 2000).

Li et al, 2002 created Barhl1 null mice by swapping the Barhl1-coding region for the *lacZ* reporter gene. In the developing mouse inner ear the Barhl1 mRNA transcript can first be detected at E14.5 in the sensory hair cells (a time by which many hair cells are already formed) and expression continues into postnatal stages. In adult mice heterozygous and homozygous for the Barhl1 null allele, expression of lacZ (which represents endogenous Barhl1 expression) suggests that Barhl1 expression continues into adulthood, particularly in the OHCs where strong β-galactosidase staining is observed. The consequence of the Barhl1 deletion in the *Barhl1*-null mice (-/-) is a profound deafness by 10 months of age that initially starts as a preferential loss of low-frequency sounds. At the cellular level, this is due to a progressive degeneration of inner and outer hair cells in the organ of Corti that is exacerbated with advancing age. In-line with the initial low-frequency loss, OHCs degenerate along an apical to basal gradient that is clearly underway by P19. IHCs, in contrast do not start to degenerate until sometime after 6 months of age and when they do so it is along the reverse gradient compared to OHCs. IHC degeneration is clearly evident at 10 months of age. Barhl1 does not appear to be involved in the initial generation, or differentiation of

inner ear hair cells; hair cells are present in the cochlea of Barhl1 null mice in early postnatal life and label for Brn-3c along with several other markers specific to hair cells. In addition, the stereociliary bundle can be observed on the hair cells of Barhl1-null mice (Li et al, 2002). Barhl1 therefore joins Brn-3c as another gene that appears paramount for the survival of cochlear hair cells. In fact, Barhl1 may be the stronger candidate for long-term maintenance of cochlea hair cells; hair cell defects in *Barhl1*-null mice exhibit a much later onset compared to Brn-3c-null mice (Erkman et al, 1996, Xiang et al, 1997, Li et al, 2002). Additionally, the presence of Brn-3c protein cannot prevent the eventual degeneration of cochlea hair cells in Barhl1-null mice (Li et al, 2002 report that Brn-3c is expressed in hair cells of Barhl1-null mice at P0-P6). Given their respective expression profiles, it is tempting to speculate that Brn-3c may be involved in regulating expression of Barhl1 in inner ear sensory hair cells. (Xiang et al, 1998, Li et al, 2002).

Li et al, proposed that Barhl1 does not directly underlie the hair cell degeneration exhibited in the Barhl1 (-/-) mice. Instead, it is suggested that Barhl1 is the key to maintenance of a component/s that is expressed in hair cells of the organ of Corti along a longitudinal gradient and crucial for their long-term survival. The voltage-gated potassium channel, KCNQ4, is suggested as a good candidate having an expression pattern in IHCs and OHCs along a gradient that reflects their degeneration in Barhl1 (-/-) mice. This hypothesis is appealing as mutations in the KCNQ4 channel are already known to be responsible for a form of autosomal dominant nonsyndromic progressive deafness, DFNA2 (Kubisch et al, 1999, Coucke et al, 1999). In DFNA2, it appears that mutations in the pore region may impair the ion selectivity of the channel upsetting the K+ homeostasis in the inner ear (Kubisch et al, 1999, Van Hauwe et al, 2000). One disparity with these observations, is that defects in KCNQ4 tend to manifest in a preferential high-frequency hearing loss, and although this is in agreement with the nature of late onset hearing loss exhibited by a large proportion of the ageing population, it is not in agreement with the nature of the hearing loss initially exhibited in the Barhl1 (-/-) mice. However, this may be due to the species difference; mice do not have the same potential as humans to accumulate exogenous damage to hair cells owing to their short lifespan. Barhl1 has yet to be identified as the defective gene in any form of human deafness, but this may just be a matter of time. In any case, Barhl1 represents a very good candidate gene for late onset hearing loss and it is possible that common sequence variants in this gene may prevail in the human population and underlie susceptibility to this disease.

1.7.4 Evidence that genes involved in ROS regulation are good candidates for late onset hearing loss.

In recent years much evidence has accumulated to implicate ROS generation in the molecular pathogenesis of late onset hearing loss (McFadden et al, 1999a; McFadden et al 1999b; Ohlemiller et al, 1999; Staecker et al, 2001). It is proposed that with ageing increased ROS generation compromises the energy status of the mitochondria in the sensory hair cells, which eventually contributes to the loss of these cells with age (for reviews see Fischel-Ghodsian et al, 2004, Ohlemiller, 2004, Pickles, 2004). In addition, there is increasing evidence that ROS generation is involved in hair cell damage following acquired environmental insults to the auditory system including acoustic trauma (Ohinata et al, 2003; Yamashita et al, 2004; for review see Lynch and Kil, 2005) and use of ototoxic agents (for reviews see Kopke et al, 1999; Campbell, 2003; Rybak and Whitworth, 2005). Accordingly, it can be reasoned that common sequence variants that lead to dysfunction or a reduction in expression levels of genes involved in ROS homeostasis have the potential to influence ones vulnerability to late onset hearing loss.

The key intrinsic antioxidant enzymes in the cochlea and other tissues that protect cells from oxidative stress include the superoxide dismutases (SODs), the glutathione peroxidases (Gpxs), catalase and the enzymes involved in glutathione regulation (for reviews see Forsberg et al, 2001; Zelko et al, 2002). Of these some of the most promising candidates for susceptibility to late onset hearing loss are SOD1 (also known as copper/zinc SOD, Cu/ZnSOD) and glutathione peroxidase-1 (Gpx1). As discussed previously (section 1.5.4) SOD1 knockout mice exhibit increased susceptibility to ARHL (McFadden et al, 1999a; McFadden et al, 1999b) and NIHL (Ohlemiller et al, 1999; McFadden et al, 2001). These findings clearly implicate SOD1 as a candidate gene for late onset hearing loss susceptibility in humans. Moreover, they have implications for the effects of ageing and acoustic trauma on late onset hearing loss pathology by suggesting a common mechanism concerning superoxide radical generation is involved. Similarly, Gpx1 like SOD1 is expressed in the cytosol and functions alongside SOD1 to regulate levels of ROS (for review see Forsberg et al, 2001). Mice with targeted deletion of the Gpx1 gene exhibit increased susceptibility to NIHL (Ohlemiller et al, 2000; McFadden et al, 2001) and exhibit histological defects characteristic of ARHL (Ohlemiller et al, 2000). Accordingly, both these factors make Gpx1 a good candidate gene for susceptibility to late onset hearing loss in humans.

Recently, a case-control association study on noise-exposed workers from an aircraft factory has revealed an association between NIHL and sequence variants in two

cellular antioxidant ezymes: SOD2 (also known as manganese SOD, MnSOD) and PON2 (paraoxonase 2) suggesting that these sequence variants predispose to NIHL (Fortunato et al, 2004). Although these associations were based on a very limited number of samples (94 noise-exposed workers of which 63 subjects were characterised with NIHL and 31 with normal hearing) and this study clearly requires replication in a larger cohort for validation; it can certainly be envisaged how polymorphisms in the *SOD2* gene may predispose to late onset hearing loss. SOD2 is localised to the mitochondria and is the sole enzyme in the mitochondria that catalyses the dismutation of the superoxide radical to hydrogen peroxide and water (for review see Zelko et al, 2002). Given that the energy status of the mitochondria is thought to be imperative for maintenance of hearing into old age (the cochlea is a highly active structure; OHCs amplify the sound input and this is a process, which requires great energy expenditure) it can certainly be perceived how any sequence variants that impaired the function or expression of *SOD2* could predispose to late onset hearing loss.

More recently, a significant association has been reported between ARHL and a sequence variant in the N-acetyltransferase 2 (*NAT2*) gene; a phase II detoxification enzyme involved in the metabolism of ROS (Unal et al, 2005). However, similar to many of the current association studies into late onset hearing loss susceptibility that have been reported in recent years (Fortunato et al, 2004; Yang et al, 2006), the study by Unal et al was based on a small number of samples (68 subjects with ARHL and 98 controls) and its validity to late onset hearing loss susceptibility awaits replication in a larger cohort.

In summary, it is clear that many genes have emerged in recent years as good candidates for susceptibility to late onset hearing loss in humans and no doubt many more will surely continue to emerge. Amongst those identified as good candidates in recent years are genes involved in stereocilia integrity (*CDH23*), hair cell survival (*BARHL1*), maintenance of calcium homeostasis in the cochlea (*PMCA2*) and regulation of ROS homeostasis in the cochlea (*SOD1, SOD2, GPX1*). Positive associations have been reported for *CDH23* (Yang et al, 2006), *SOD2* and *PON2* (Fortunato et al, 2004) and susceptibility to NIHL and between *NAT2* and susceptibility to ARHL (Unal et al, 2005). However, each of these studies requires confirmation in a larger cohort for validation. In addition, in the absence of functional analysis it is not clear whether the sequence variants reported by Fortunato et al, and Yang et al are the causal mutations that are be responsible for the hearing loss pathophysiology.

At the onset of this project *Brn-3c* was one of the most promising candidate genes for susceptibility to late onset hearing and at present, still remains so. *Brn-3c* as a candidate gene for late onset hearing loss is explored in detail in the next section, 1.8.

1.8 The candidate gene approach: Brn-3c.

At the onset of this project there was a scarcity of candidate genes for susceptibility to late onset hearing loss; *Brn-3c*, a POU domain transcription factor was one of the most promising candidates. *Brn-3c* is one of three highly homologous genes collectively known as the Brn-3 family categorised under division IV of the widely characterised POU protein family (for review see Latchman, 1999). *Brn-3a* (also known as *Brn3.0, POU4F1*) and *Brn-3b* (also known as *Brn-3.2, POU4F2*) the additional members of the Brn-3 family are important developmental regulators involved in the maintenance and survival of specific neuronal cell sub-types in the nervous system and retina, respectively (Erkman et al, 1996; Gan et al, 1996; McEvilly et al, 1996; Xiang et al, 1996; Eng et al, 2001 and discussed further in section 1.9.2). In recent years much evidence has accumulated to suggest that Brn-3c functions in maturation and maintenance of sensory hair cells during development and ultimately as a life-long pro-survival factor for these hair cells (Erkman et al, 1996, Xiang et al, 1997, Xiang et al, 1998 and Vahava et al, 1998). This is important as loss of auditory hair cells appears to be the main cause of late onset hearing loss (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller et al, 2004; Gates and Mills, 2005). Evidence in support of the proposal that *Brn-3c* is a good candidate gene for late onset hearing loss will now be discussed.

1.8.1 Evidence to implicate Brn-3c as a candidate gene for late onset hearing loss.

In the nervous system and certain regions of the brain Brn-3c is expressed in a specific but overlapping expression pattern with other members of the Brn-3 family (for review see Ryan and Rosenfeld, 1997). However, in the inner ear Brn-3c is the only Brn-3 factor expressed in hair cells and expression is confined to the sensory hair cells of both the auditory and vestibular systems (Erkman et al, 1996, Xiang et al, 1997, 1998). Immunolabeling studies in mice with a specific anti-Brn-3c antibody show that in the otocyst Brn-3c expression initiates early during development at E12.5 in presumptive hair cells of the developing sensory epithelia (Xiang et al, 1998). Strong and selective expression of Brn-3c can then be clearly detected in auditory and vestibular hair cells from embryonic day 14.5, a time by which time all inner ear sensory epithelia are formed (Xiang et al, 1997, 1998). Expression appears restricted to postmitotic hair cells

(Xiang et al, 1998) and importantly is maintained postnatally at high levels into adulthood (Erkman et al, 1996, Xiang et al, 1997, 1998).

The functional significance of specific Brn-3c expression in inner ear sensory hair cells is evident in Brn-3c 'knockout mice' (-/-) which were generated independently by two laboratories through targeted deletion of the Brn-3c gene (Erkman et al, 1996, Xiang et al, 1997). Brn-3c (-/-) mice exhibit complete loss of cochlear and vestibular hair cells by early postnatal life followed by severe degeneration of spiral and vestibular ganglia neurons (Erkman et al, 1996, Xiang et al, 1997). This is because although hair cells initially generate in Brn-3c (-/-) mice and undergo the initial stages of differentiation they fail to undergo full maturation. Cells morphologically similar to hair cells can be observed in the organ of Corti of Brn-3c (-/-) mice at E16.5 which immunostain positively for hair cell specific markers myosin VI and VIIa (Xiang et al, 1998). However, the ability to acquire complete specialization is abolished; the molecular cues needed for formation of the highly organized stereociliary hair bundle appear lost. Indeed, the apical surface of the organ of Corti is completely clear of stereociliary hair bundles by P0 (Erkman et al, 1996, Xiang et al, 1997, 1998). Furthermore, in the vestibular system partially differentiated hair cells fail to undergo correct migration (Xiang et al, 1998). The consequence of absence of Brn-3c on the morphology of the inner ear is specific; in adult Brn-3c (-/-) mice the exterior portion, middle ear, and overall infrastructure of the cochlear and vestibular system remain intact (Xiang et al, 1997). In the cochlea aberrations are only seen in the neighbouring areas alongside the loss of the inner and outer hair cells; the columnar Pillar cells characteristic of the tunnel of Corti are generally absent and Deiter's cells cannot be clearly defined (Xiang et al, 1997). Terminal dUTP nick end labelling (TUNEL) on serial sections of sensory epithelia from Brn-3c (-/-) mice shows that the incompletely differentiated hair cells progressively degenerate by apoptosis during development such that the total loss of these cells is evident in the sensory epithelia shortly after birth (Xiang et al, 1998). The loss of the spiral and vestibular ganglia neurons in the Brn-3c (-/-) mice also proceeds by apoptosis (Xiang et al, 1998), but is regarded a secondary effect of hair cell loss because Brn-3c is restricted to hair cells in the inner ear. It is proposed that the sensory ganglia neurons depend on the hair cells for trophic support. Indeed, hair cells are known to express brain-derived neurotrophic factor (BDNF) and neurotrophin-3 (NT-3) and sensory neurons express the appropriate neurotrophic receptors (for reviews see Fritzsch and Beisel, 2001, Marzella and Gillespie, 2002).

Moreover, recent in-vitro evidence suggests that Brn-3c may play a role in BDNF and NT-3 expression in hair cells (Clough et al, 2004).

Taken together, the evidence from Brn-3c (-/-) mice shows that presence of Brn-3c protein is needed for hair cell maturation and maintenance. It is feasible that as a POU domain transcription factor Brn-3c regulates expression of genes needed by hair cells for complete maturation and survival. Certainly, expression of the Brn-3c protein is needed for formation of the highly organized stereociliary hair bundle (Erkman et al, 1996, Xiang et al, 1997, 1998). In addition, it is clear that the complete loss of Brn-3c, at least in mice, will ultimately end in hair cell death; directly or indirectly presence of Brn-3c protein can prevent the progressive apoptotic degeneration of inner ear sensory hair cells (Xiang et al, 1998). Hair cells are paramount to auditory and vestibular function. Consequently the morphological defects in the Brn-3c (-/-) mice manifest physically as complete deafness coupled with balance and co-ordination defects and marked hyperactivity (Erkman et al, 1996, Xiang et al, 1997). Interestingly, this phenotype appears limited to mice homozygous for targeted deletion of Brn-3c, mice that are heterozygous and express one functional copy of the Brn-3c gene are indistinguishable from the wild type; they appear normal in all aspects (Erkman et al, 1996, Xiang et al, 1997).

Significant evidence in agreement with the notion that *Brn-3c* is a good candidate gene for late onset hearing loss in humans is obtained from a study of hearing loss within a large Israeli-Jewish family (Vahava et al, 1998). In this family, an eight base-pair deletion in the POU homeodomain of *Brn-3c* causes adult onset hearing loss DFNA15 with an autosomal dominant mode of progressive late onset hearing loss (Vahava et al, 1998). In affected family members the hearing loss phenotype manifests between the ages 18 to 30 years and progresses to a moderate to severe deafness by 50 years of age. All affected family members tested heterozygous for the 8bp deletion (Vahava et al, 1998). Hence, in this family at least, possession of one wild type allele of *Brn-3c* is not sufficient to maintain normal hearing past early adulthood. The fact that the hearing loss is not congenital and exhibits a delayed onset suggests that Brn-3c is needed to maintain hair cell function in the long-term. Taken together with the evidence from Brn-3c (-/-) mice (Erkman et al, 1996, Xiang et al, 1997 and 1998) this implies that Brn-3c may function as a pro-survival factor for inner ear sensory hair cells not only during development but also throughout life. Indeed, Brn-3c is ideally placed to function as a life-long pro-survival factor for hair cells. Brn-3c expression within the inner ear is restricted to hair cells and is maintained into adulthood (Erkman et al, 1996,

Xiang et al, 1997 and 1998). Furthermore, as a POU domain transcription factor and not a cytoskeletal component of the hair cells, Brn-3c has the ability to modulate expression of many genes needed for hair cell function and survival. These features are very appealing when considering *Brn-3c* as a candidate gene for late onset hearing loss because loss of auditory hair cells appears to be the main cause of this disease.

1.8.2 Understanding the molecular mechanism by which an 8bp deletion in the POU homeodomain of Brn-3c causes adult onset hearing loss, DFNA15.

In the Israeli-Jewish family with DFNA15 the presence of one wild-type allele of *Brn-3c* is not sufficient to maintain hearing past early adulthood (Vahava et al, 1998). Two proposals that have been suggested to explain the mechanism of the hearing loss in family members affected by DFNA15 are haploinsufficiency or a dominant negative effect of the mutant Brn-3c protein on the function of the wild-type Brn-3c protein (Vahava et al, 1998). In DFNA15 the 8bp deletion is in exon 2 of the *Brn-3c* gene and this is predicted to result in a frame shift commencing at codon 295 and inducing premature termination of translation at codon 299. It is thought that in members of the Israeli family with DFNA15, a truncated protein of 298 amino acids is produced in addition to the 338 amino acid wild-type protein. This concept is supported by *in-vitro* analysis; the mutated transcript is translated in cell culture producing a truncated protein of the predicted size relative to the wild-type protein (Weiss et al, 2003).

Given that the mutated transcript is probably translated in members of the Israeli family with DFNA15 if haploinsufficiency of *Brn-3c* (half the normal levels of Brn-3c) underlies the hearing loss phenotype then this is probably through functional redundancy of the mutant copy of *Brn-3c*. Haploinsufficiency resulting from functional redundancy of the mutant Brn-3c protein, a truncated protein of 298 amino acids is certainly feasible. The DNA binding affinity of the mutant protein is predicted to be low or completely abolished because the deletion removes the majority of the carboxyl terminus of the POU homeodomain including the recognition helix that makes base contacts with the DNA (Vahava et al, 1998). Recent in-vitro experiments aimed to delineate the effect of the 8bp deletion on the function of the Brn-3c protein are in agreement with this notion; *in-vitro* translated mutant Brn-3c exhibits minimal binding to a consensus Brn-3 binding site (Weiss et al, 2003). If mutant Brn-3c protein has reduced affinity for its DNA recognition motifs *in-vivo* then expression of genes

involved in hair cell survival will certainly be compromised. Furthermore, impaired DNA binding ability may not be the only factor leading to functional redundancy of mutant Brn-3c protein. Two nuclear localisation signals have been identified in the Brn-3c protein: a monopartite element spanning amino-acids 274 to 278 and a bipartite element spanning amino-acids 314 to 331 (Weiss et al, 2003). In members of the Israeli family that carry the 8bp deletion it is predicted that the bipartite nuclear localisation signal is lost. The consequence of loss of the bipartite nuclear localisation signal on the function of the Brn-3c protein is impaired nuclear targeting; immunolocalisation analysis shows that around 50% of the mutant protein is retained in the cytoplasm (Weiss et al, 2003). Hence, Brn-3c-dependent expression of hair cell survival genes could be further compromised by the cytoplasmic compartmentalisation of a normally nuclear transcription factor. In addition, the first amino acid lost in the 8bp deletion is isoleucine at position 22 of the POU homeodomain (Vahava et al, 1998). This position has been established as an area of functional significance in the Brn-3 family (Dawson et al, 1996b, Dawson et al, 1998, Budhram-Mahadeo et al, 1998) and its loss in the mutant Brn-3c protein is likely to have a profound effect on the function of the protein. Indeed, data obtained from transient transfection assays are in agreement with the concept that expression of downstream targets of Brn-3c are compromised by the nature of the 8bp deletion; the mutant Brn-3c protein, truncated at 298 amino-acids loses most of its transactivating ability (Weiss et al, 2003, Clough et al, 2004).

Interestingly, mice that lack one functional copy of Brn-3c do not develop adult onset hearing loss; Brn-3c (+/-) mice at the end of their life-span do not show a greater hearing loss or exhibit a greater extent of cochlear degeneration compared to equivalent aged-wild-type animals (Keithley et al, 1999). This is at odds with the notion that haploinsufficiency of *Brn-3c* underlies adult onset hearing loss in DFNA15. However, using the mouse as a model organism for late onset hearing loss in humans has its limitations. The data reported by Keithley et al, 1999 could be limited by the mouse strain used in the study, but more importantly mice have a short lifespan living to three years at most and the mice used by Keithley et al, 1999 were raised in a controlled environment. This contrasts widely with the situation for humans. Humans live much longer, are exposed to environmental insults and accordingly have the potential to accumulate exogenous damage to hair cells over a very long period of time. In patients with DFNA15 if hair cells are already 'vulnerable' at the molecular level due to less Brn-3c protein product they may only be able to sustain a certain level of exogenous distress before the onset of hearing loss. The late onset of hearing loss phenotype seen

in DFNA15 is consistent with this suggestion; individuals heterozygous for the 8bp deletion do not exhibit hearing loss until 18-30 years of age. Furthermore, it should be considered that Brn-3c expression could be subject to autoregulation in mice and that this compensates for the loss of one allele in Brn-3c (+/-) mice by up-regulating transcription from the single copy allele so levels of Brn-3c protein are not dissimilar to wild-type. This is certainly possible. Autoregulation of gene expression is a common control mechanism amongst eukaryotic transcription factors (for review see Bateman, 1998), many members of the POU transcription factor family appear subject to autoregulation (Delhase et al, 1996; Malik et al, 1996; Trieu et al, 2003) and in Brn-3a (+/-) mice autoregulation of Brn-3a has been shown to compensate for the loss of one allele of Brn-3a (Trieu et al, 2003). Nevertheless to explain the different phenotype in humans and mice with one functional copy of Brn-3c Vahava et al originally proposed that mutant Brn-3c in DFNA15 may impair the functioning of the wild-type Brn-3c protein by a dominant negative effect (Vahava et al, 1998). However, when tested experimentally mutant Brn-3c protein shows little ability to bind to a consensus sequence for the Brn-3 family and does not interfere with the binding of wild-type Brn-3c (or Brn-3b) to this sequence (Weiss et al, 2003). In agreement with this observation, mutant Brn-3c loses most of its transactivating ability and co-transfections of wild-type and mutant Brn-3c expression constructs show that mutant Brn-3c protein does not affect the ability of wild-type Brn-3c protein to transactivate luciferase reporter gene constructs containing downstream targets of Brn-3c (Weiss et al, 2003, Clough et al, 2004). These experiments suggest that a dominant negative effect of the mutant Brn-3c protein is unlikely to be responsible for the hearing loss phenotype in members of the Israeli family with DFNA15. However, it must be considered there are many mechanisms by which a mutant protein can exert a dominant negative effect over the function of the wild-type protein. Accordingly it cannot be ruled out that *in-vivo* mutant Brn-3c hinders contact with a co-factor of wild-type Brn-3c. The N-terminal region of mutant Brn-3c protein is probably intact *in-vivo*; this could compete with wild-type Brn-3c for binding to an as yet unidentified protein target of Brn-3c required for some Brn-3c regulation.

Finally it should also be considered that the hearing loss phenotype in members of the Israeli family that carry the 8bp deletion in Brn-3c may be due to a 'gain of function' mechanism acquired by the mutant Brn-3c protein; mutant Brn-3c may have novel functional properties. This is certainly possible, *in-vitro* analysis shows that the 298 amino acid mutant Brn-3c protein loses a bipartite nuclear localisation signal and

consequently around 50% of the mutant protein is retained in the cytoplasm (Weiss et al, 2003). In addition, pulse-chase analysis has shown that the mutant Brn-3c protein is more stable than the wild-type protein (Weiss et al, 2003). These novel features acquired by the mutant Brn-3c protein maybe deleterious to hair cell function over time.

In summary, it is clear that an 8bp deletion in the POU homeodomain of *Brn-3c* causes adult onset hearing loss, DFNA15. *In-vitro* analysis with mutant Brn-3c protein shows that its DNA binding ability and transcriptional activity are all but lost, nuclear targeting is impaired and protein stability of the mutant protein is enhanced compared to wild type (Weiss et al, 2003, Clough et al, 2004). Weiss et al, propose that it is this repertoire of functional defects that ultimately ends in cochlear hair cell loss and adult onset hearing loss, DFNA15. The exact mechanism by which the 8bp deletion in *Brn-3c* results in adult onset hearing loss is speculative, but there is strong evidence to suggest functional redundancy of mutant Brn-3c. Consequently, haploinsufficiency of *Brn-3c* may compromise expression of genes involved in hair cell survival. This is important, as if haploinsufficiency of *Brn-3c* underlies DFNA15 and is sufficient to induce adult onset hearing loss manifesting between ages 18-30 years it can certainly be envisaged how mild sequence variants in *Brn-3c* might contribute to late onset hearing loss in the ageing population. For example, a common sequence variant in the promoter of *Brn-3c* could have a subtle effect on *Brn-3c* expression that is not phenotypically evident until the fifth or sixth decade of life. Certainly, a dominant negative effect of the mutant Brn-3c protein on the function of the wild-type Brn-3c protein appears unlikely. A gain of function mechanism is possible but currently there is no evidence to suggest the mutant Brn-3c protein has novel functional properties that are deleterious to hair cells. Finally, it should also be considered that the mechanism by which the 8bp deletion in *Brn-3c* causes adult onset hearing loss DFNA15 may not be attributed to a single mechanism but rather a mixture of molecular mechanisms that manifest in the late onset hearing loss phenotype. The 8bp deletion in the POU homeodomain of *Brn-3c* identified by Vahava et al, has only been identified in one family, the Israeli-Jewish family. It appears to be a rare mutation; it has not been shown to underlie hearing loss in any other families. Evidence as discussed here suggests that Brn-3c is paramount for hair cell maturation and long-term survival. Hence, it is conceivable how common sequence variants in the *Brn-3c* gene that have a moderate effect on function or expression of this gene may be a risk factor for late onset hearing loss that is exhibited by a large proportion of the ageing population.

1.9 The Brn-3 family: the wider picture.

The POU domain transcription factor Brn-3c is, as has been discussed, clearly a good candidate gene for susceptibility to late onset hearing loss in humans. Brn-3c together with Brn-3a and Brn-3b, two highly homologous genes, constitutes the Brn-3 family (for review see Latchman, 1999) the founder member of which, Brn-3a, was originally identified in 1989 by He et al. The background to the Brn-3 family will now be explored followed by molecular and functional characterization of this family with special emphasis placed on Brn-3c and how these findings relate to the role of Brn-3c in inner ear sensory hair cells.

The sequence homology shared between the Brn-3 POU factors extends to the I-POU factor from *Drosophila* and the Unc-86 factor from the nematode *C. elegans*. Collectively these POU proteins form a subset division IV, of a much larger and widely characterized family of POU proteins. Currently on the basis of sequence homology up to six subsets of POU factors exist (for details of the other subsets see the following reviews: Verrijzer and Van der Vliet, 1993; Andersen and Rosenfeld, 2001). The term "POU" refers to the DNA-binding domain shared by members of the POU protein family and derives from the original four family members shown to share this region of homology: Pit-1/GHF-1, Oct-1, Oct-2, and Unc-86. Pit-1 belongs to subset I of the POU protein family; expression is restricted to the anterior pituitary gland where it is essential in development and maintenance of specific cell types (for review see Andersen and Rosenfeld, 2001). The octamer binding transcription factors Oct-1 and Oct-2 are categorized under subset II. Oct-1 is ubiquitously expressed and is important in regulation of housekeeping genes. In contrast, Oct-2 has a more restricted expression pattern confined to B-lymphocytes of the immune system and specific neuronal cells of the nervous system where it has been shown to play an important role in regulation of gene expression (for review see Latchman et al, 1999). Unc-86, the first *C.elegans* POU domain protein to be identified (Finney et al, 1998) belongs to subset IV and is a critical regulator of neuronal cell lineage and terminal differentiation of neurons (for review see Ryan and Rosenfeld, 1997). It is well known that POU proteins function as developmental regulators. They are important in development of the nervous, endocrine and immune system where they play important roles in determination of cell lineage, cell proliferation, migration, terminal differentiation and survival in these systems (for reviews see Wegner et al, 1993, Latchman et al, 1999 and Andersen and Rosenfeld, 2001).

1.9.1 The POU domain.

The POU domain is the evolutionary conserved DNA binding domain shared by members of the POU protein family. It is a bipartite structural entity typically stretching 150-160 amino acids and consisting of an amino-terminal POU specific domain (POU_S) and a carboxyl-terminal POU homeodomain (POU_H). The POU_S is a highly conserved moiety that is unique to the POU protein family and is composed of four α-helices spanning 74-82 amino-acids. A conserved cluster of basic amino acids defines the first and fourth helices and the boundaries of the POU_S are enriched in acidic residues. The POU_H is the second most highly conserved region and is related to that found in the classic homeobox proteins. The POU_H comprises three α-helices over a 60 amino acid stretch with clusters of basic amino-acids characterising the boundaries of the POU_H. In each sub-domain the second and third helix form the distinctive helix-turn-helix motif, and the third helix, the recognition helix, is able to make base contacts in the major groove of DNA (for reviews see Verrijzer and Van der Vliet, 1993; Andersen and Rosenfeld, 2001).

A variable linker region of 15-27 amino acids separates each sub-domain. The linker plays an important role in conferring flexibility upon DNA binding (For review see Phillips and Luisi, 2000). A linker length between 10-14 amino acids facilitates binding of the class II POU factor Oct-1 to an octamer site whereas shorter linker lengths place constraints on the DNA binding ability of Oct-1 (Van Leeuwen et al, 1997). Both the POU_S and POU_H are important for DNA binding; cooperative binding of the sub-domains allows efficient sequence-specific recognition of asymmetrical DNA-binding motifs that are typical for POU domain proteins (Aurora et al, 1992).

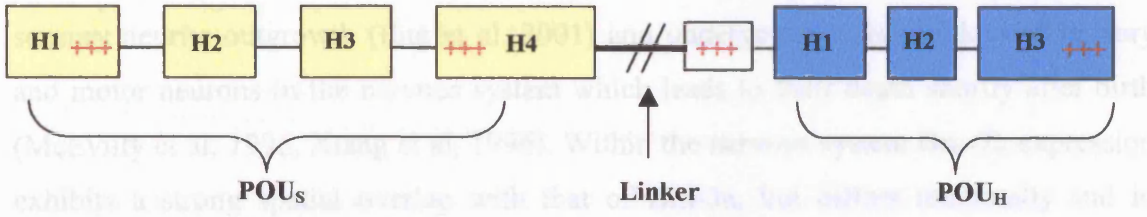


Figure 1.5 A schematic diagram to illustrate the bipartite POU domain. Diagram is not to sacle.The POU domain is characterised by an amino-terminal POU specific domain (POU_S) and a carboxyl-terminal POU homeodomain (POU_H). The POU_S domain consists of four α-helicies (shown as yellow shaded boxes H1 – H4) the first and fourth of which are defined by a region of basic amino acids (shown as +++ in diagram). The POU_H contains three α-helicies (shown as blue shaded boxes H1 – H3); a region of basic amino acids precedes helix 1 and defines the carboxyl boundary of helix 3. A variable linker region (indicated by the solid black line with two hatches) separates each sub-domain. The function of each sub-domain and the linker is discussed within the text.
Reproduced and modified from Wegner et al, 1993.

Many POU proteins have been reported to bind DNA as monomers, but some are capable of binding as homo- and hetero-dimers (for reviews see Phillips and Luisi, 2000; Andersen and Rosenfeld, 2001). It is not clear how Brn-3 factors bind to their cognate sites *in-vivo*; although dimer formation has been observed on an experimentally derived consensus sequence for the Brn-3 family (Gruber et al, 1997). The function of the POU domain is not limited to DNA binding; it is also capable of mediating specific protein-protein contacts that are important for functional activity of POU domain proteins (for review see Andersen and Rosenfeld, 2001).

Strong sequence homology in the POU domain exists between the Brn-3 family members; in humans they share approximately 95% sequence identity at the amino acid level (Xiang et al, 1995). Of the 11 amino-acid residues in the POU domain that are exchanged between Brn-3 family members, 6 reside in the linker region (Xiang et al, 1995). Upstream of the POU domain, similarity between the Brn-3 factors is less well conserved. The exception is a N-terminal region termed the upstream homology domain (UHD) that is also found in the other class IV POU domain factors (Xiang et al, 1995).

1.9.2 Molecular and functional characterisation of the Brn-3 family: with emphasis on Brn-3c.

The Brn-3 factors share distinct but overlapping patterns of expression in the nervous system and are all involved in the terminal differentiation and survival of specific cell types. Brn-3a is the most widely expressed factor of the Brn-3 family, expression ranges from peripheral sensory neurons to specific nuclei of the central nervous system where it is plays an important role in the differentiation and survival of neuronal cell subtypes. Mice with targeted deletion of the Brn-3a gene exemplify this, they exhibit defects in sensory neurite outgrowth (Eng et al, 2001) and undergo an extensive loss of sensory and motor neurons in the nervous system which leads to their death shortly after birth (McEvilly et al, 1996, Xiang et al, 1996). Within the nervous system Brn-3b expression exhibits a strong spatial overlap with that of Brn-3a, but differs temporally and in relative protein abundance. This is particularly evident within the ganglion cell layer of the retina where the onset of Brn-3b expression commences several days earlier than that of Brn-3a (for review see Ryan and Rosenfeld, 1997). Accordingly, mice with targeted deletion of the Brn-3b gene exhibit a specific loss of retinal ganglion cells leading to blindness and do not die after birth (Erkman et al, 1996, Gan et al, 1996). Subsequent analysis has established that Brn-3b is not involved in initial fate

determination of retinal ganglion cells but plays a paramount role in differentiation, axonal outgrowth and survival of these cells (Gan et al, 1999, Wang et al, 2000). Brn-3c expression is the most restricted of the Brn-3 family; expression in the nervous system is limited (for review see Ryan and Rosenfeld, 1997), it is expressed in the retina but is the last Brn-3 factor to be expressed (Xiang et al, 1995) and Brn-3c is the only Brn-3 factor expressed in hair cells of the inner ear (Erkman et al, 1996, Xiang et al, 1997, 1998). The significance of this expression pattern is reflected in the Brn-3c (-/-) mice as discussed previously (see section 1.8.1); these mice are viable, but deaf with balance defects due to loss of sensory hair cells from the vestibulo-cochlea system of the inner ear (Erkman et al, 1996, Xiang et al, 1997).

The Brn-3 family is encoded by separate genes on separate chromosomes but exhibits closeness in sequence homology. In humans, Brn-3c has been mapped to chromosome 5q31 (Xiang et al, 1995), Brn-3a to chromosome 13q21.1-q22 (Xiang et al, 1995) and Brn-3b to chromosome 4q31.2 (Xiang et al, 1993). The human Brn-3 factors are closely related to their murine counterparts exhibiting 98% amino-acid identity (Xiang et al, 1995). Mouse orthologues of Brn-3a, 3b and 3c have been mapped to mouse chromosomes 14E1-E3, 8F1-F5 and 18B3-E1, respectively (Theil et al, 1994). Structurally, the human Brn-3 gene family are very similar composed of two small exons intervened by a small intron at the same relative position in the protein coding region (Xiang et al, 1995) and this structural similarity extends to their mouse homologues (Theil et al, 1993, 1994). Brn-3a is well characterised at the molecular level; short [Brn-3a(S)] and long [Brn-3a(L)] isoforms have been identified in human (Xiang et al, 1995) and mice (Thiel et al, 1993) which appear to be under the control of two distinct promoters for the Brn-3a gene (Frass et al, 2002, Thomas et al, 2004). Both isoforms contain the C-terminal POU domain but the short form lacks an 84 amino-acid motif termed the N-terminal activation domain (Thiel et al, 1993). To date no short isoform of Brn-3c has been identified only a long form of 388 amino-acids (Xiang et al, 1995) which, similar to Brn-3a, is characterised by an N-terminal activation domain (Smith et al, 1998b; Sud et al, 2005) in addition to the POU domain at the C-terminus.

Currently, the mechanism by which Brn-3c functions as a POU domain transcription factor and regulates gene expression in sensory hair cells of the inner ear is not well understood. This contrasts with the situation for Brn-3a in neuronal cells where much more progress has been made. The distinct domains of Brn-3a are well characterised and allow Brn-3a to function in a bi-functional manner (for review see Latchman et al, 1999, see also Faulkes et al, 2004). In neuronal cells Brn-3a(L) has a

well established role in preventing apoptosis (for review see Latchman, 1999), at least in part via the activation of Bcl-2 (Smith et al, 1998a) and Bcl-x (Smith et al, 2001) and these effects are dependent upon the N-terminal activation domain of Brn-3a. Consequently, the Brn-3a(S) which lacks the N-terminal activation domain is limited in function; it cannot enhance neuronal survival. However, both forms of Brn-3a (Brn-3a(L) and Brn-3a(S)) can activate expression of genes involved in neurite outgrowth because in contrast to Brn-3a(L) activation of Bcl-2 and Bcl-x these effects are dependent upon the POU domain of Brn-3a functioning as both a DNA binding and transactivation domain; the valine residue at position 22 of the POU_H being paramount for this effect (for review see Latchman et al, 1999).

In neuronal cell lines it has been shown that Brn-3c is able to transactivate the same genes as Brn-3a, although generally, to a weaker extent (Budhram-Mahadeo et al, 1995, Smith et al, 1998b). Co-transfections in ND-7 cells of CAT reporter construct carrying either the SNAP-25, α-internexin, or one of the three neurofilament subunit promoters with a Brn-3c expression vector showed Brn-3c could activate each of these promoter constructs previously shown to be activated by Brn-3a (Smith et al, 1998b). This is conceivable as the Brn-3c factors are capable of binding to the same binding sites although subtle differences in affinity for binding site sequence may exist (Gruber et al, 1997). However, in the absence of *in-vivo* data it is questionable whether these neuronally expressed genes are genuine targets of Brn-3c.

Smith et al, 1998b, reported that the transactivating effect of Brn-3c was limited to neuronal cell lines and in contrast to Brn-3a-dependent activation of the same promoter constructs could not be achieved by the POU domain in isolation. Indeed, by deleting discrete regions of Brn-3c and creating Brn-3c fusion proteins with the GAL-4 DNA binding domain the trans-activating ability of Brn-3c was localised to a 177 amino-acid region in the N-terminal of Brn-3c. Since the transactivating ability of intact Brn-3c and the Brn-3c N-terminal activation domain as a GAL4 fusion protein was only observed in neuronal cell lines, it was concluded by Smith et al. that the Brn-3c N-terminal activation domain is neuronal cell specific and a neuronal cell specific co-activator was suggested to be responsible for this effect. However, recent evidence disputes the notion that the N-terminal activation domain of Brn-3c is neuronal cell specific (Clough et al, 2004, Sud et al, 2005). It has recently emerged that Brn-3c can transactivate specific BDNF and NT-3 promoter constructs in inner ear sensory epithelial derived cell lines, OC-1 and OC-2, and this is dependent on functional Brn-3c with an intact N-terminal domain (Clough et al, 2004). Hence, suggesting that at least

one mechanism whereby Brn-3c may function in inner ear sensory hair cells is by transactivating the expression of downstream targets via its N-terminal activation domain. Recent molecular analysis on Brn-3c by Sud et al, 2005 is in agreement with this suggestion. In an attempt to delineate the mechanism by which Brn-3c functions in hair cells Sud et al, 2005 linked discrete regions of Brn-3c to the DNA binding domain of the heterologous transcription factor GAL4 (similar to Smith et al, 1998b in neuronal cells) and assayed the ability of these constructs to regulate the activity of a promoter containing GAL4 DNA binding sites in OC-1 and OC-2 cell lines. The results of this analysis showed that only the intact N-terminal domain of Brn-3c, expressed as a GAL4 fusion protein, is able to strongly activate luciferase reporter constructs containing Gal-4 binding sites. Sub-fragments of the Brn-3c N-terminal expressed as Gal-4 fusion proteins, or a construct expressing the isolated POU domain were not able to achieve this effect. Furthermore, overexpression of the Brn-3c N-terminal activation domain as a discrete moiety did not interfere with the activation of the GAL4 heterologous promoter by the Brn-3c:GAL4 fusion protein containing the intact N-terminal activation domain of Brn-3c. Hence, suggesting that the ability of the Brn-3c N-terminal domain to function as an activation domain in inner ear sensory hair cells is due to an intrinsic property of this domain and not the recruitment of a co-factor.

The failure of the isolated POU domain of Brn-3c to function as a dual activating and DNA binding domain on promoter constructs shown to be activated by Brn-3c in neuronal cell lines (Smith et al, 1998b) and inner ear sensory epithelial derived cell lines (Clough et al, 2004, Sud et al, 2005) may be due to the fact that the POU_H domain of Brn-3c contains the amino acid isoleucine at position 22. This position is a critical determinant of functional activity amongst members of the Brn-3 family (Dawson et al, 1996b, 1998). It is responsible for Brn-3a and Brn-3b functioning as an antagonistic pair of transcription factors (for review see Latchman et al, 1999). Indeed, Brn-3a generally functions to transactivate target promoters for example, the genes involved in neurite outgrowth (SNAP-25 and the three neurofilament subunits) whereas, Brn-3b represses expression of these genes and also prevents their activation by Brn-3a (for review see Latchman et al, 1999). These opposing behaviours of Brn-3a and Brn-3b relate to the valine present at position 22 of the POU_H in Brn-3a and the isoleucine found at this position in Brn-3b, and are suggested to be mediated by specific protein-protein interactions at this position (Dawson et al, 1998). The POU domain of Brn-3c containing the same amino-acid as Brn-3b at position 22 of the POU_H, isoleucine, is

hence unlikely to function as a discrete activating-DNA binding domain in inner ear sensory hair cells.

Despite a clear difference at position 22 of the POU_H, the POU domain is almost identical between members of the Brn-3 family (Xiang et al, 1995) and the Brn-3 factors exhibit highly similar DNA binding properties on consensus sequences derived by random oligonucleotide selection (Xiang et al, 1995, Gruber et al, 1997). This has lead to the suggestion that the Brn-3 factors may mediate downstream events via a shared molecular mechanism and hence, be functionally equivalent. Recently, some evidence has been gathered in agreement with this suggestion (Wang et al, 2002, Pan et al, 2005). Studies from Brn-3b/Brn-3c double knockout mice suggest that Brn-3c can partially compensate for loss of Brn-3b in the retina; Brn-3b/Brn-3c double knockout mice exhibit a greater loss of retinal ganglion cell axons compared to Brn-3b knockout mice alone (Wang et al, 2002). Similarly, but using a knock-in approach Brn-3a has been shown to compensate functionally for Brn-3b in retinal ganglion cells (Pan et al, 2005). However, this finding is at odds with the antagonist mechanism of action observed for Brn-3a and Brn-3b in neuronal cells (for review see Latchman et al, 1999). As Brn-3c is the only Brn-3 factor expressed in inner ear sensory hair cells, it is possible that targeted replacement of the Brn-3c gene with that of Brn-3a, or Brn-3b may help clarify if the mechanism by which Brn-3c functions in these cells involves a common pathway to that of Brn-3a or Brn-3b in neuronal or retinal ganglion cells, respectively.

In summary, evidence suggests that Brn-3c functions as a transcriptional activator in sensory hair cells of the inner ear; the N-terminal activation domain of Brn-3c being crucial for this effect. It is also possible that due to containing isoleucine at position 22 of the POU_H Brn-3c may function as a repressor of gene transcription and hence, contain dual activating / repressing properties although, this remains to be confirmed. In any case, it is clear that given the selective expression pattern of Brn-3c in hair cells, the evidence gained from Brn-3c (-/-) mice and the evidence for the role of other members of the Brn-3 family in controlling gene expression in other cell types, Brn-3c is likely to be a critical determinant of coordinating gene expression in hair cells. Furthermore, these genes are likely to be involved in hair cell maintenance and survival – attributes that make Brn-3c an excellent candidate gene for susceptibility to late onset hearing loss.

1.9.3 How does Brn-3c act as a pro-survival factor for hair cells?

It is clear, from *in-vitro* molecular analysis that Brn-3c may function as a transcriptional activator in inner ear sensory hair cells. In addition, the presence of isoleucine at position 22 of the POU_H raises the possibility that Brn-3c may also function as a transcriptional repressor. Taken together, this suggests that Brn-3c may be a bi-functional transcription factor in inner ear sensory hair cells. However, currently little is known about bona fide target genes of Brn-3c in hair cells or in other tissues within the nervous system where Brn-3c is expressed and how they are regulated. Progress has been hindered by the extremely limited number of hair cells per cochlea (only about 16,000 in mammals) and the difficulties in accessing these cells and maintaining them in culture. However, it is evident from patients with DFNA15 and studies with Brn-3c (-/-) mice that downstream targets of Brn-3c are likely to be involved in hair cell maturation and maintenance. Indeed there is evidence that Brn-3c may be directly involved in regulating neurotrophin gene expression in the inner ear, expression of which is crucial for innervation of the mature hair cell (Clough et al, 2004).

From studies with Brn-3c (-/-) mice it is clear that loss of Brn-3c leads to hair cell death; directly or indirectly presence of Brn-3c protein can prevent the progressive apoptotic degeneration of hair cells (Xiang et al, 1998). There is increasing evidence from studies with rodents to suggest apoptosis underlies the aetiology of hair cell loss associated with ageing and development of late onset hearing loss (Usami et al, 1997, Zheng et al, 1998, Alma et al, 2001; and discussed in section 1.5.4). Therefore, it is possible that at least one mechanism whereby Brn-3c functions as a long-term pro-survival factor for hair cells is via continued involvement in apoptosis. The Bcl-2 family is a well-characterised family of proteins involved in apoptosis in many cell types (for review Chan and Yu, 2004). Interestingly, as discussed, Brn-3a, a close homologue of Brn-3c that functions as a survival factor for neuronal cells can induce the expression of anti-apoptotic members of this family in particular, Bcl-2 (for review of the evidence see Latchman, 1998) and Bcl-x(L) (Smith et al, 2001). Hence, it is very tempting to speculate that Brn-3c functions in hair cells in a manner not dissimilar to that of Brn-3a in neuronal cells, namely as an anti-apoptotic factor by directly regulating expression of genes that are members of the apoptosis cascade. The molecular mechanisms by which Bcl-2 functions as an anti-apoptotic factor are complex and have yet to be fully elucidated but one such mechanism appears to be via inhibiting the generation of reactive oxygen species from the mitochondria (for review see Jang and Surh, 2003).

Recently a spontaneous mouse mutant, dreidel (ddl), has been identified in Brn-3c (see http://jaxmice.jax.org/strain/003484.html). The ddl mutation removes a TG dinucleotide at 338bp in the Brn-3c coding region and is predicted to result in a truncated protein lacking the POU domain. The (ddl/ddl) mice are deaf and the hearing loss phenotype is reported to be similar to that of the Brn-3c (-/-) mice leading the suggestion that the (ddl/ddl) mice hold promise as a second mouse model for loss of function of Brn-3c. By comparing mRNA expression profiles from E16.5 inner ears of (ddl/ddl) and wild-type mice, the transcription factor growth factor independence 1 (Gfi-1) has been reported as a downstream target gene of Brn-3c in hair cells (Hertzano et al, 2004). Hertzano et al, reported similarities between outer hair cell morphology and degeneration in Brn-3c (-/-), (ddl/ddl) and Gfi-1 (-/-) mice and suggested that the loss of outer hair cells in the Brn-3c (-/-) and (ddl/ddl) mice is due to a failure of Brn-3c to maintain Gfi-1 expression. However, more evidence is needed to clarify this issue. The evidence provided by Hertzano et al, 2004 does not establish whether Brn-3c regulates Gfi-1 directly and morphological comparisons were only made at certain developmental time points. Hence, although it is possible that Gfi could underlie some of the outer hair cell loss seen in the Brn-3c (-/-) and (ddl/ddl) mice it is presumptive to conclude that it underlies 'all' the outer hair cell loss observed. Furthermore, in Gfi-1 (-/-) mice inner hair cells are present along the entire length of the cochlea duct, at least at E18.5, and this contrasts widely with the situation for Brn-3c mutant mice. Nevertheless the work of Hertzano et al, does raise the possibility that Brn-3c may have an indirect effect on members of the apoptosis cascade, as at least one role of Gfi-1 that has been characterised in other cell types, is to prevent apoptosis. Gfi-1 can inhibit cell death by directly repressing the pro-apoptotic gene Bax (Grimes et al, 1996b). Additionally, as highlighted by Hertzano et al, 2004 Gfi-1 can promote STAT3 signals by a virtue of its ability to sequester the STAT3 inhibitor PIAS3 (Rodel et al, 2000), and STAT3 signalling has been reported to transactivate members of the anti-apoptotic Bcl-2 family (Fukada et al, 1996; Catlett-Falcone et al, 1999). The identification of STAT3 protein in the outer hair cells of the cochlea by Hertzano et al, is in agreement with this concept.

One well-known stimulus whereby a cell's fate can be pushed towards apoptosis is withdrawal of growth factors or when survival signals become limiting. With this in mind it is of interest that growth factors have been shown to maintain mammalian hair cells *in-vitro*; OHC losses in adult quinea pig organ of Corti explants can be prevented by treatment with many growth factors including acidic fibroblast growth factor, insulin-like growth factor-1 and epidermal growth factor amongst others (Malgrange et

al, 2002). It is tempting to speculate that Brn-3c may function as a long-term pro-survival factor for hair cells by regulating expression of growth factors needed for their maintenance. Interestingly, patients with Turner syndrome that is characterized by a lack of the growth factor estrogen, commonly undergo an early presbyacusis (Hultcrantz et al, 1994) suggesting that the growth factor estrogen may be important in maintenance of hearing. There appear to be very few studies reported in the literature regarding estrogen and hearing although it has been suggested that estrogens may confer a protective effect on hearing (Jonsson et al, 1998). Estrogen receptors are reported to be expressed in auditory hair cells of mice and rats (Stenberg et al, 1999), and in the 'Turner mouse' an animal model of Turner syndrome, premature presbyacusis is observed concomitant with deterioration of outer hair cells in the cochlea (Hultcrantz et al, 2000). Interestingly, there is evidence that members of the Brn-3 family, Brn-3a and Brn-3b can functionally interact with the estrogen receptor (Budhram-Mahadeo et al, 1998), but whether this applies to Brn-3c and has significance to the pro-survival role of Brn-3c in hair cells can only be speculated upon at present.

In summary, the downstream target genes of Brn-3c in inner ear sensory hair cells are largely unidentified at present. However, it is clear from evidence gained with Brn-3c (-/-) mice and patients with DFNA15 that downstream targets of Brn-3c are likely to function in hair cell maturation and long-term survival. Currently, experiments are underway using both *in-vitro* (our laboratory) and *in-vivo* (Avraham et al) approaches to identify Brn-3c target genes. It is hoped that by utilising both these approaches and identifying downstream targets of Brn-3c in hair cells elucidation of hair cell survival pathways will be facilitated. By gaining a greater understanding of the molecular players in hair cell survival therapeutic intervention to slow and ultimately prevent loss of hearing with advancing years will be aided.

1.10 The aim of the project.

The aim of this project is to investigate genetic susceptibility to late onset sensorineural hearing loss (as defined in section 1.5.1) using a candidate gene case-control association study approach. The candidate gene to be investigated is *Brn-3c* (also known as *Brn3.1* and *POU4F3*), a POU domain transcription factor that within the inner ear is specifically expressed in hair cells (Erkman et al, 1996; Xiang et al, 1997, 1998). At the onset of this project *Brn-3c* was one of the most promising candidate genes for susceptibility to late onset hearing loss. In recent years, many additional promising candidate genes for susceptibility to late onset hearing loss have emerged (reviewed in section 1.7). However, *Brn-3c* still remains one of the strongest candidates. Collectively, evidence suggests that Brn-3c functions in hair cell maturation and maintenance during development and ultimately as a life-long pro-survival factor for hair cells (Erkman et al, 1996, Xiang et al, 1997, Xiang et al, 1998 and Vahava et al, 1998). Given that loss of auditory hair cells appears to be the main cause of late onset hearing loss (for reviews see Jennings and Jones, 2001, Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills, 2005) these features make *Brn-3c* an excellent candidate gene for susceptibility to this disease (a detailed discussion of the evidence to implicate *Brn-3c* as a good candidate gene for late onset hearing loss is reviewed in section 1.8).

Late onset hearing loss is an extremely common and complex disease. Genetic predisposition to late onset hearing loss, as with all traits of a complex nature, is thought to be due to moderate variations in gene expression or function at multiple loci. The CD:CV hypothesis states that genetic susceptibility to common disease is due to common genetic variants present with a rare allele frequency of or greater than 1% in the general population (for reviews see Smith and Lusis, 2002; Wright et al, 2003). Therefore, in line with the CD:CV model it was hypothesized that common inter-individual sequence variants in the *Brn-3c* gene or within the regulatory regions of this gene that have a subtle affect on function or expression of Brn-3c, respectively could be a risk factor for susceptibility to this disease. To investigate this hypothesis four main aims were devised for this project that are summarised and discussed below:

1) To identify common sequence variants in the *Brn-3c* gene.

2) To use *in-vitro* assays where available and assess whether common sequence variants in the *Brn-3c* gene are functional.

3) To initiate collection of a well-characterised patient cohort for late onset hearing loss.

4) To assess whether common and functional sequence variants in the *Brn-3c* gene are a risk factor for late onset hearing loss by means of a pilot case-control association study.

1.10.1 Identification of common sequence variants in the *Brn-3c* gene.

Common sequence variants are proposed to underlie susceptibility to common disease and are reported to exist with a rare allele frequency of or greater than 1% in the general population; this is the basis of the CD:CV hypothesis (for reviews see Smith and Lusis, 2002; Wright et al, 2003). At the time of commencing this project minimal information was reported in the public SNP databases for the *Brn-3c* gene. Therefore, in an attempt to identify common sequence variants in the *Brn-3c* gene that may predispose to late onset hearing loss it was decided to perform mutation scanning of the *Brn-3c* gene in a small sample consisting of members of the general population. This line of analysis is explored in Chapter 4.

1.10.2 Establishing whether common sequence variants in the *Brn-3c* gene are functional.

If common sequence variants are identified in the *Brn-3c* gene these could be 'causal' and pathogenic for late onset hearing loss or they could be neutral variants that have no affect on susceptibility to this disease. As discussed previously (section 1.6.2 and 1.6.3) use of functional sequence variants in case-control association analysis is extremely important. Use of well-characterised functional sequence variants increases the chances of finding a positive association as the likelihood of a false-positive association due to identification of what is in effect a neutral sequence variant in linkage disequilibrium with the casual variant at the same or another locus is reduced. In addition, it is only

when the function and the effect(s) of the causal sequence variant have been functionally characterised that its role in disease pathogenesis can be well understood and its effect on disease risk can be evaluated with confidence. Hence, making for more robust interpretation of case-control association data and increasing the likelihood of replication in subsequent independent studies. Therefore, for the purpose of this project it was decided that any common sequence variants identified in the *Brn-3c* gene would be assayed for functional impact with respect to *Brn-3c* by appropriate *in-vitro* analysis. For example, common sequence variants within the promoter of *Brn-3*c could result in subtle inter-individual differences in *Brn-3c* expression by modifying cis-acting elements and hence, the binding affinity of transcription factors. Therefore, if common sequence variants within the promoter of *Brn-3c* are identified in this project these will be tested in-vitro to see if they modify transcription factor binding by using nuclear protein extracts from a suitable cell line. Common promoter variants that have an affect on protein binding will then be assessed further by utilising suitable cell lines for reporter gene assays to determine if differences in sequence and hence binding affinity manifest in functional differences in regulation of the *Brn-3c* gene. Similarly, if common sequence variants are identified in the coding region of *Brn-3c* or the intron (the *Brn-3c* gene is characterised by a single non-coding intron) these too will be tested *in-vitro* for functional impact using suitable techniques as appropriate depending on the location of the sequence variant. The rationale behind techniques used for functional analysis is discussed further in Chapter 5 and functional analysis of common sequence variants in the *Brn-3c* gene is discussed throughout Chapters 5 - 7.

1.10.3 Collection of a well-characterised patient cohort for late onset hearing loss.

The long-term aim of this project is to assess whether common and functional sequence variants in the *Brn-3c* gene are a risk factor for late onset hearing loss by means of a case-control association study (see next section, 1.10.4). In order for this part of the aim to stand a chance of success large and well-characterised patient cohorts are needed. However, at the onset of this project well-characterised patient cohorts for late onset hearing loss were scarce. Consequently, it was an important aim of this project to initiate collection of a large, well-characterised patient cohort for late onset hearing loss that could be used in subsequent association analysis. Therefore, in accord with this aim sample collection from subjects characterised by a consultant audiologist as having a

late onset hearing loss was initiated from the adult hearing aid clinic at the Royal Free Nose, Ear and Throat Hospital, London, U.K. at the onset of this project. Ethical approval for this part of the project was granted from the Royal Free Local Research Ethics Committee (6202). More information concerning this aspect of the project is described in Method section 3.2.1 and discussed in Chapter 8 section 8.2.

1.10.4 Establishing whether common and functional sequence variants in the *Brn-3c* gene are a risk factor for late onset hearing loss by means of a case-control association study.

The ultimate aim of this project is to use a case-control association study approach to assess whether common and functional sequence variants in the *Brn-3c* gene are a risk factor for late onset hearing loss. As has been discussed previously (section 1.5) late onset hearing loss is a complex disease and consequently, the relative risk contributed by any one sequence variant is thought to be small. Hence, in order for the power of the association study to be sufficient to detect a positive association large and well-characterised patient cohorts are needed. Collection of a late onset hearing loss patient cohort was initiated (see section 1.10.3) in order to perform association analysis in accord with this part of the aim. However, it was acknowledged at the outset of this project that the number of patients with late onset hearing loss that it would be possible to recruit into this study within the time frame of this PhD project may limit the power of the association analysis that could be performed. Therefore, the association analysis performed for this PhD project was considered a preliminary study. This line of analysis is discussed further in Chapter 8.

In summary, by fulfilling these aims genetic variation in the *Brn-3c* gene would be characterised and through functional analysis insight may be gained into how Brn-3c an important hair cell pro-survival factor functions and / or how this gene is regulated. Indeed, currently the molecular mechanism by which Brn-3c functions as a POU domain transcription factor in inner ear sensory hair cells is not well understood and very little is known as to how the *Brn-3c* gene is regulated; the downstream targets of Brn-3c are largely unidentified and no transcription factors involved in regulation of Brn-3c have been elucidated. By collection of a well-characterised patient cohort for late onset hearing loss one would be able to evaluate with confidence the affect of Brn-3c as a risk factor for late onset hearing loss by using sequence variants only for which there is evidence that they are functional in subsequent association analysis. It was

accepted at the outset of this project that this line of analysis would be limited by the power of the preliminary association study that it would be possible to perform within the time constraints of this PhD project. However, a preliminary association study would provide information on allele frequency and pave way for a large-scale population based case-control association study, which is beyond the time frame of this PhD project. It is through performing this type of analysis that a greater understanding of the genetic factors that underlie susceptibility to late onset hearing loss should be achieved.

1.11 Publications before submission of this thesis.

Before submission of this thesis two important publications have come to light:

Johnson et al, 2006. Strain background effects and genetic modifiers of hearing in mice. Brain Res. 1091: 79-88.
Report five additional Ahl loci (Ahl4 – Ahl8) that have been mapped in mice.

Van Laer et al, 2006. The contribution of genes involved in potassium-recycling in the inner ear to noise-induced hearing loss. Hum Mutat. 27: 786-795.
Report an association between sequence variants in three genes (KCNE1, KCNQ1 and KCNQ4) involved in potassium recycling in the inner ear and susceptibility to noise induced hearing loss. The authors report the sequence variant p.D85N in KCNE1 as a possible causative sequence variant based on subsequent functional analysis (this is discussed further in section 9.0: General Discussion and Conclusion).

2.0 The Aim Of The Project.

The aim of this project is to investigate *Brn-3c* as a candidate gene for susceptibility to late onset hearing loss. To investigate this hypothesis four main aims were devised for this project that are summarised below and discussed in detail in Chapter 1, section 1.10:

1) To identify common sequence variants in the *Brn-3c* gene.

2) To use *in-vitro* assays where available and assess whether common sequence variants in the *Brn-3c* gene are functional.

3) To initiate collection of a well-characterised patient cohort for late onset hearing loss.

4) To assess whether common and functional sequence variants in the *Brn-3c* gene are a risk factor for late onset hearing loss by means of a pilot case-control association study.

3.0 Materials And Methods

3.1 Materials.

3.1.1 Chemicals and equipment.

Chemicals were purchased from Amersham Biosciences, BDH, Invitrogen, Promega and Sigma unless otherwise stated.

Restriction endonucleases were purchased from Promega.

[γ-^{32} P]-ATP and [α-^{32}P]-dCTP were purchased from Amersham Biosciences.

Nunclon™ plasticware for tissue culture was purchased from Fisher Scientific.

Experimental manipulations were performed using autoclaved MilliQ double distilled water (ddH_2O).

3.1.2 Plasmids.

pGEM®-T easy vector, pGL3-Basic vector, pGL3-Promoter vector, pRL-SV40 vector, phRL-null vector, pSi mammalian expression vector and pCi mammalian expression vector were purchased from Promega.

The mammalian expression vector pEVR2-Sp1 was kindly provided by Professor Guntram Suske, University of Marburg, Germany and has been described previously (Hagen et al, 1992).

The *Drosophilia* expression vector pPac-USp3 was kindly provided by Professor G.Suske as before and has been described previously (Dennig et al, 1996).

The mammalian dominant negative SP1 expression vector pEBGN-SP1 was kindly provided by Professor Gerald Thiel, University of the Saarland Medical Centre, Germany and has been described previously (Petersohn and Thiel, 1996; Lietz et al, 2003).

The mammalian expression vector CMV5-Gfi-1 was kindly provided by Professor Horwitz, University of Washington and has been described previously (Grimes et al, 1996b).

3.1.3 Abbreviations.

ATP	Adenosine 5'-triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DMSO	Dimethyl sulphoxide
dNTPs	Dideoxynucleotide triphosphates
dTTP	deoxythymidine triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra-acetic acid
EGTA	Ethylene glycol-bis (β-aminoethyl ether)-N,N,N'N'-tetraacetic acid.
IPTG	Isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani medium
NP40	Nonidet-P-40
OC-2	organ of Corti-2
PBS	Phosphate buffered saline
rpm	revolutions per minute
SDS	Sodium dodecyl sulphate
SSC	Sodium chloride/sodium citrate buffer
TAE	Tris-actate-EDTA electrophoresis buffer
TBE	Tris-borate-EDTA electrophoresis buffer
TEMED	N,N,N'N'-tetramethylethylenediamine
X-gal	5-bromo-4chloro-3-indolyl-β -D-glactropyranoside

3.1.4 Stock solutions.

Stock solutions were made up using autoclaved MilliQ deionised water unless otherwise stated.

6x DNA loading buffer:	15% Ficoll, 0.25% bromophenol blue, 0.25% xylene cyanol.
Denaturing solution:	1.2M NaCl, 0.4M NaOH.
Hybridisation solution:	6x SSC, 0.5% SDS, 2.5x Denhardts solution.
100x Denhardts solution:	2% Ficoll, 2% Polyvinyl Pyrrolidone, 2% BSA.
SDS-polyacrylamide gel	
5% Stacking gel:	5% acrylamide, 0.125M Tris (pH 6.8), 0.1% SDS, 0.1% ammonium persulfate, 0.1% TEMED.
8% Resolving gel:	8% acrylamide, 0.375M Tris (pH 8.8), 0.1% SDS, 0.1% ammonium persulfate, 0.06% TEMED.
1x SDS gel-loading buffer:	50mM Tris-Cl (pH 6.8), 2% SDS, 10% glycerol, 0.1% bromophenol blue, 100mM DTT.
10x Western running buffer:	0.25M Tris, 1.92M glycine, 1% SDS.
Blotting buffer:	25mM Tris, 192mM glycine, 30% methanol.
Tris Buffered Saline-Tween: (TBS-T)	20mM Tris (pH 7.6), 150mM NaCl, 0.1% Tween-20.
TBS-T (10% non fat milk powder):	10% (w/v) non-fat milk powder (Marvel), TBS-T.
TBS-T (5% non fat milk powder):	5% (w/v) non-fat milk powder (Marvel), TBS-T.

Formamide gel-loading buffer:	98% formamide, 0.1% xylene cyanol, 0.1% bromophenol blue, 9.8mM EDTA (pH 8.0).
4% non-denaturing polyacrylamide gel (for EMSA analysis):	4% acrylamide, 0.25x TBE, 0.037% TEMED, 0.07% ammonium persulphate.
10x TBE:	0.89M Tris-borate, 20mM EDTA (pH 8.0).
50x TAE:	2.0M Tris, 1M Glacial acetic acid, 50mM EDTA (pH 8.0).
Parker buffer 10μM Zinc:	20mM Hepes (pH 7.9), 50mM KCl, 0.5mM DTT, 8% ficoll, 10μM zinc sulphate.
Parker buffer 100μM Zinc:	20mM Hepes (pH 7.9), 50mM KCl, 0.5mM DTT, 8% ficoll, 100μM zinc sulphate.
L-Broth:	1% tryptone (DUCHEFA), 0.5% yeast extract (DUCHEFA), 0.17M NaCl.
LB-Agar:	L-broth containing 1.5% Micro Agar (DUCHEFA).
20x SSC:	3M NaCl, 0.3M trisodium citrate.

3.2 Methods.

3.2.1 Patient samples

For initial mutation scanning of the Brn-3c gene genomic DNA samples were obtained from 45 random individuals and kindly provided by Dr Maria Bitner-Glindzicz, Institute of Child Health, University College London, U.K. Genomic DNA samples were obtained from individuals predominantly of Caucasian ethnicity.

For the association study patients were characterised by Dr Barbara Cadge, Audiological Physician, Institute of Laryngology and Otology, University College London, U.K as having a late onset sensorineural hearing loss (for definition see section 1.5.1) and were recruited from the adult hearing aid clinic at the Royal Free Nose, Ear and Throat Hospital, London, U.K. Ethical approval was granted from the Royal Free Local Research Ethics Committee (6202). Patients were identified in advance from the forthcoming clinic appointments lists and sent the Patient Information Sheet, Patient Questionnaire: Adult Onset Hearing Loss and Consent form (see Appendix A). Where available identification of suitable patients was based on diagnostic audiogram. Patients were excluded from recruitment if: less than 30 years of age, history of known dementia and / or learning disability or patient identified as requiring an interpreter (which are likely to cause problems with the consent procedure). On day of attendance at clinic all patients in receipt of the Patient Information Sheet were asked if they wished to accept or decline the invitation to participate in the study. If accepted, signed consent was obtained, patients were asked to donate a blood sample (5mls) and data sets were collected on patients according to the Patient Questionnaire (see Appendix A). Data sets collected on patients included gender, age, ethnicity, diagnostic audiogram, age of onset, family history of hearing loss, history of noise exposure including both acute (sudden) and chronic exposure (for example, occupational noise exposure) in addition to details of significant medical problems. Upon review of patient data sets, patients were further excluded from the study if: they were non-Caucasians, upon diagnostic audiogram hearing loss was identified as conductive, a specific known aetiology of the sensorineural hearing loss was identified for example: acoustic neuroma (a non-cancerous growth near auditory nerves), Meniere's disease, evidence of head trauma leading to sudden sensorineural hearing loss or evidence of acute noise exposure leading to sudden sensorineural hearing loss. Patients were not excluded from the study if there was an asymmetric hearing loss as long as the better hearing ear was consistent with the criteria for late onset sensorineural hearing loss. In these cases onset of hearing loss was

dated from the better hearing ear characteristic of late onset sensorineural hearing loss and the worst hearing ear (due to hearing loss manifesting from childhood infections, mastoid surgery or Meniere's disease) was ignored.

For the control sample group genomic DNA samples were obtained from random, healthy members of the general population and kindly provided by Professor Patricia Woo, Centre for Paediatric Adolescent Rheumatology, Windeyer Institute, University College London, U.K. and have been described previously (see Crawley et al, 1999). Data sets available on subjects included gender, age and ethnicity. For the association study subjects of Caucasian ethnicity were selected.

3.2.2 The polymerase chain reaction (PCR).

All PCR reactions were performed with an Eppendorf Mastercycler Gradient.

3.2.2.1 PCR for SSCP analysis.

Primers were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to generate overlapping PCR amplicons covering the complete coding region of Brn-3c (NCBI GenBank Accessions: U10060.1 and U10061.1), the intron (NCBI GenBank Accession: AF043452.1), and 1.5Kb of the 5'-flanking region (Reference sequence: Ensembl transcript ID ENST00000230732; see Appendix B). An additional set of primers were designed to span a dinucleotide $(GT)_n$ repeat at -3495bp within the 5'-flanking region. Primers were purchased from Qiagen (http://oligos.qiagen.com) and re-suspended at 100μM in ddH_2O; sequence of primers is summarised in Table 3.1, section 3.2.18. Each PCR reaction was optimised to ensure specific amplification of the required band by varying the magnesium concentration and annealing temperature of the reaction and visualisation of a single band of the correct size by agarose gel electrophoresis. Reactions were performed on 100ng genomic DNA, with 0.2μM of each primer, 200μM dNTPs, 10mM Tris-HCL (pH9.0), 50mM KCl, 0.1% Triton X-100, 2.0mM $MgCl_2$, and 0.5U of *Taq* DNA polymerase (Promega) in a total volume of 50μl. Reactions were supplemented with DMSO 5% (v/v) as required (see Table 3.1, section 3.2.18). Thermal cycling was performed under the following conditions: denaturation at 94°C for 30 seconds, annealing at optimal pre-determined temperature for 30 seconds, extension at 72 °C for 30 seconds, for 30 cycles, with an initial denaturation step of 94°C for 5 minutes and a final extension of 72°C for 7 minutes.

3.2.2.2 PCR for site-directed mutagenesis at -1391A>C.

PCR was performed on 10ng double-stranded plasmid DNA (p-Luc-1391C) using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene) with 125ng each primer: -1391aSDM-S (5' GCAGCGTAGTCGAGGTCCAGGATTC 3') and -1391tSDM-AS (5' GAATCCTGGACCTCGACTACGCTGC 3'), 1µl of dNTP mix, 1x reaction buffer and 2.5U of *PfuTurbo* DNA polymerase mix in a total volume of 50µl. Thermal cycling was performed at 95°C for 30 seconds followed by 12 cycles of: 95°C for 30 seconds, 55°C for 1 minute and 68°C for 8 minutes. Post-PCR 40µl of the PCR mix was used in a *DpnI* digest to digest the template plasmid DNA (non-mutated) and 1µl of the *DpnI*-treated DNA was used in a transformation reaction according to the manufacturer's instructions (the target nucleotide at -1391 for site-directed mutagenesis is highlighted in black bold within the primer sequence shown).

3.2.2.3 PCR for generation of a 710bp fragment spanning -566$(GT)_n$ in the Brn-3c 5'-flanking region.

To generate a 710bp fragment spanning -566$(GT)_n$ in the Brn-3c 5'-flanking region for use in sub-cloning PCR was performed on 100ng genomic DNA of pre-determined genotype at -566$(GT)_n$ based on the initial SSCP screen. Primers 3cP4-S and 3cP10-AS (for sequence see Table 3.1, section 3.2.18) were used in a standard PCR reaction as described in section 3.2.2.1 but with an annealing temperature of 60.3°C, an annealing and an extension stage of 45 seconds and PCR reactions were supplemented with 5% DMSO.

3.2.2.4 PCR for generation of a 398bp fragment spanning -3495$(GT)_n$, -3457$(GA)_n$ and -3432poly-G polymorphism in the Brn-3c 5'-flanking region.

To generate a 398bp fragment spanning -3495$(GT)_n$, -3457$(GA)_n$ and -3432poly-G polymorphism in the Brn-3c 5'-flanking region PCR was performed on 100ng genomic DNA with primers 3cP12-S: 5' GAGCTCCTGAAGCAGTGTCT 3' and
3cP12-AS: 5' GCAATTTGTAACAGCCTCTATGC 3' in a standard PCR reaction as described in section 3.2.2.1 but with an annealing temperature of 59°C. Post-PCR the resulting amplicon was sub-cloned into pGEM®-T easy vector (Promega) for sequencing.

3.2.2.5 PCR for construction of expression vector pSi-DBD-SP1.

PCR was performed on 100ng double-stranded plasmid DNA (pEBGN-SP1) using the BD Advantage ™ 2 PCR Kit (BD Biosciences) for high fidelity PCR with 0.2μM each primer: Sp1DomNeg-S (5' ATGCTGGTTCCGCGTGGATCAC 3') and Sp1DomNeg-AS (5' CCCTCACTCTAGAGTCGC 3'), 200μM dNTPs, 1x Advantage 2 PCR buffer, 1x Advantage 2 DNA polymerase mix in a total volume of 50μl. Thermal cycling was performed at 95°C for 1 minute followed by 30 cycles of: 95°C for 30 seconds and 68°C for 1 minute, with a final extension of 70°C for 7 minutes.

3.2.3 Single strand conformational polymorphism (SSCP) analysis.

All amplicons generated by PCR for SSCP analysis were within the size range 130 - 302bp (see Table 3.1, section 3.2.18). For SSCP, a 2μl aliquot of each amplicon was mixed with 6μl of Formamide reagent (52% Formamide, 10mM EDTA, 0.05% SDS, 0.02 % Bromophenol blue and 0.02 % Xylene cyanol). Samples were heated at 90°C for 3 minutes, snap-cooled on ice and loaded onto a non-denaturing polyacrylamide gel with pre-formed sample wells (CleanGel 48S, DNA Analysis Kit, Pharmacia Biotech. Dimensions after rehydration 250 x 110 x 0.5mm, stacking gel T = 5 %, C = 3 %, separating gel T = 10 %, C = 2 %). SSCP was performed on a Multiphor II electrophoresis unit with a Multitemp thermostatic circulator unit at 4, 10 and 20 °C using the CleanGel DNA analysis kit (Pharmacia Biotech). Bands were visualised by silver staining according to the DNA Silver Staining Kit (Pharmacia Biotech). Amplicons that displayed abnormal SSCP migration patterns indicative of a sequence variant were directly cloned into the pGEM®-T easy vector (Promega) for sequencing. Both strands were sequenced on an ABI Prism 3100 Genetic Analyser (see section 3.2.7 and Table 3.3, section 3.2.18) and compared to the reference sequence (see Appendix B). Variants identified are described with respect to the first A of the ATG translation start as +1.

3.2.4 Cell culture.

3.2.4.1 Culture of OC-2 cells.

The mouse inner ear sensory epithelial cell line, Organ of Corti-2 (OC-2) was obtained from Professor Matthew Holley, University of Bristol, U.K.. OC-2 cells were

maintained at 33°C, 5 % carbon dioxide in Eagle's minimal essential medium (Biowhittaker) supplemented with glutamax (2mM), 10% foetal calf serum and 50U/ml γ-interferon. Cells were seeded at a density of approximately 1-2 $x10^5$ cells per $75cm^2$ flask (Nunclon™) and passaged when cells reached approximately 80% confluency.

3.2.4.2 Preparation of OC-2 nuclear extracts.

Nuclear extracts were prepared from OC-2 cells using an adaptation of Alkalay et al, 1995. Briefly, OC-2 cells were grown in large $175cm^2$ flasks (Nunclon™) until 80 % confluent, washed once in cold PBS (calcium and magnesium free) and covered in 3mls of cold cytoplasmic lysis buffer (10mM Hepes pH 7.6, 1mM EDTA, 0.1mM EGTA, 10mM KCl, 1mM DTT, 1mM sodium pyrophosphate, 1mM sodium orthovanadate and 1x Complete, Mini protease inhibitor cocktail, Roche). Incubated on ice for 10 minutes, harvested with a rubber policeman and transferred to a series of eppendorf tubes. NP40 was added to the top side of each eppendorf at a final concentration of 0.6%, tubes were vortexed for a few seconds (left no longer than 5 minutes) and extracts were spun down at 13,000 rpm for 30 seconds at 4 °C. The supernatant was removed and all traces of cytoplamsic lysis buffer removed from the pellet. 50µl of nuclear lysis buffer (20mM Hepes pH 7.6, 0.2mM EDTA, 0.1mM EGTA, 25% glycerol, 0.42M NaCl, 1mM DTT, 1mM sodium pyrophosphate, 1mM sodium orthovanadate and 1x Complete, Mini protease inhibitor cocktail, Roche) was added to each pellet and nuclear proteins were extracting by freeze-thawing by transferring eppendorfs from an ethanol/dry ice bath to a 37°C water bath and repeating three times. Extracts were spun down at 13,000 rpm for 10 minutes at 4 °C and the supernatant containing nuclear proteins, pooled. The protein concentration was determined by spectroscopy using the formula below* and extracts were aliquoted, snap frozen and stored at -80°C.

*Concentration $\mu g/\mu l = (1.55 \times A_{280}) - (0.76 \times A_{260})$.

3.2.4.3 Preparation of OC-2 nuclear extracts transfected with an expression vector for Gfi-1.

The day before transfection OC-2 cells were plated into 90mm dishes (Nunclon™) at 5 x 10^5 cells per dish. OC-2 cells were transfected with 1.2µg of either pCi expression vector (Promega) or CMV5-Gfi-1 expression vector according to the calcium phosphate precipitation method (Gorman, 1985); three dishes of transfected cells were prepared for each expression vector. Twenty-four hours after transfection the medium was removed

and a standard glycerol shock was performed (15% glycerol). Forty-eight hours after glycerol shock cells were harvested for nuclear extract as described in section 3.2.4.2 using 1ml of cold cytoplasmic lysis buffer per dish.

3.2.5 *In-Vitro* Transcription and Translation.

Generation of *in-vitro* translated (IVT) SP1 and SP3 protein was performed using pSi-SP1 and pSi-SP3 expression plasmids as templates (expression plasmids are described in section 3.2.8.2). Reactions were performed on 1μg of each expression plasmid using the TnT® T7 Quick Coupled Transcription / Translation System (Promega) according to the manufacturer's instructions.

3.2.6 The electrophoretic mobility shift assay (EMSA).

3.2.6.1 Design and annealing of oligonucleotides.

Complementary oligonucleotides were designed to span the sequence variation on wild type and variant alleles and to include any putative transcription factor binding sites as identified from using MatInspector software (www.genomatix.de/products/MatInspector/index.html). Sequence of consensus binding sites and mutated binding sites were obtained from the Santa Cruz Biotechnology, Inc. website (www.scbt.com) unless stated otherwise. Oligonucleotides are summarised in Table 3.2 (section 3.2.18); 5'- overhangs were added to each oligonucleotide as specified in Table 3.2 to facilitate subsequent sub-cloning. All oligonucleotides were purchased from Qiagen (http://oligos.qiagen.com) and resuspended at 1μg/μl in ddH_2O.

Complementary oligonucleotides were annealed by combining 10μg of each oligonucleotide with Maniatis medium salt buffer (50mM NaCl, 10mM Tris-Cl pH 7.5, 10mM $MgCl_2$ and 1mM DTT), heating at 95°C for 10 minutes and slowly cooling to room temperature. Annealed oligonucleotides were visualised as retarded bands on 3 % agarose gels stained with ethidium bromide (0.25μg/ml); annealed oligonucleotides migrate significantly slower than their single stranded counterparts.

3.2.6.2 5'end-labelling of annealed oligonucleotides.

Annealed oligonucleotides were 5' end-labelled in a standard T4 polynucleotide kinase reaction with [γ-^{32} P] ATP (Sambrook et al, 1989) to generate radioactive probes for use

in EMSA analysis; T4 polynucleotide kinase catalyses the transfer of the γ-phosphate from ATP to the 5'-terminal, hydroxyl end of polynucleotides. Briefly, 50ng of annealed oligonucleotide was combined with 3μl of [γ-^{32} P] ATP, 5U of T4 polynucleotide kinase and 1x standard kinase buffer (Promega) in a total volume of 20μl. The labelling reaction was allowed to proceed for 30 minutes to 1 hour at 37°C. After which, the reaction volume was diluted to 100μl with ddH_2O and the labelled, double stranded oligonucleotides purified by passage through a G25 Sephadex column with centrifugation (1500rpm for 3 minutes). The success of the labelling reaction for each probe was determined by measuring the counts per second (cps) with a mini-monitor Geiger counter (Mini 900 Ratemeter, Thermo Electron Corporation). Probes producing readings ≥ 2Kcps/100μl at a distance of 1cm were used in EMSA analysis.

3.2.6.3 The binding reaction.

Binding reactions and subsequent analysis was performed according to the protocol described below unless otherwise stated. Binding reactions were performed in Parker buffer (20mM Hepes pH 7.9, 50mM KCl, 1mM EDTA, 0.5mM DTT and 8 % ficoll) with 3μg of poly (dI/dC) and where appropriate either 8-10 μg of OC-2 nuclear protein extract, 129-258ng recombinant human SP1 protein (Promega) or appropriate quantity of *in-vitro* translated SP1 or SP3 protein in a total volume of 20μl. Binding reactions were pre-incubated on ice for 10 minutes before the addition of the γ-^{32} P labelled probe (1ng) and then for a further 30 minutes on ice. For competition experiments unlabelled competitor DNA was added after the 10 minute pre-incubation at 25- to 1000-fold molar excess and allowed to incubate for a further 20 minutes on ice prior to addition of labelled probe. For supershift analysis 1-3μl of: mouse monoclonal anti-SP1 antibody [SP1 (1C6) X; Santa Cruz Biotechnology, Inc.], goat polyclonal anti-Gfi-1 antibody [Gfi-1 (N-20) X; Santa Cruz Biotechnology, Inc.], rabbit polyclonal anti-OCT-2 antibody [OCT-2 (C-20) X; Santa Cruz Biotechnology, Inc] or mouse monoclonal anti-ATF-1 antibody [ATF-1 (C41-5.1) X; Santa Cruz Biotechnology, Inc.] were added to the binding reaction subsequent to addition of labelled probe and allowed to incubate for 30 minutes on ice. Samples were loaded onto a pre-run (30 minutes at 200V) 4 % non-denaturing polyacrylamide gel (dimensions: 200 x 200 x 3mm) and electrophoresed in 0.25x TBE for 2 ½ hours at 200V, at 4 °C using standard Bio-RAD electrophoretic equipment. The gel was placed on Whatman 3MM paper, covered with Saran wrap and dried with a Gel Master™ gel dryer vacuum system (Jencons-PLS, model 1426) for 2

hours at 80°C. Shifted protein-DNA complexes were visualised by autoradiography of the dried gel on BioMax MS (Kodak), BioMax MR (Kodak) or Hyperfilm™ MP (Amersham Biosciences) film as appropriate, at -80°C.

3.2.7 DNA Sequencing.

DNA sequencing was performed by Mr Ian Gerrard or Mr John Cheshire at Windeyer Sequencing, Windeyer Institute, University College London, U.K. Approximately 100ng purified PCR product or 200-500ng purified plasmid DNA was sequenced on an ABI Prism 3100 Genetic Analyser (PE Biosystems). The fluorescent dye deoxy terminator methodology that utilises the BigDye chemistries (BigDye terminator, PE Biosystems) was employed by cycle sequencing with AmpliTaq™ DNA Polymerase FS; this way, labelled ddNTP terminators are incorporated during the asymmetric PCR. *Alternatively*, DNA sequencing was performed by Mr Colin Herd at MRC geneservice, Babraham Bioincubator, Cambridge, U.K. on an ABI 3730 DNA Analyser using BigDye terminator chemistries version 3.1 (BigDye terminator, PE Biosystems). Primers used in sequencing are summarised in Table 3.3 (section 3.2.18). DNA sequencing chromatograms were visualised using BioEdit Sequence Alignment Editor software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

3.2.8 Plasmid construction.

3.2.8.1 Human Brn-3c promoter constructs.

The luciferase reporter vector pGL3-Basic (Promega) carrying 3.6Kb (bases -80bp to -3670bp) of the human Brn-3c promoter (pgl3b-Brn3c-3.6) had already been cloned in our laboratory from the human PAC library RPCI1 (HGMP, MRC, U.K.) by Mr Sanjay Jagutpal; this construct was used as template to form a series of Brn-3c promoter-luciferase reporter gene constructs that each carried a distinct 5' haplotype of the Brn-3c promoter. The genetic background of the Brn-3c promoter insert in pgl3b-Brn3c-3.6 at regions established from the initial PCR-SSCP screen to be polymorphic was determined by direct sequencing to be -3495$(GT)_{15}$, -3457$(GA)_2$, -3432$(G)_8$CG (poly-G allele SNPG1), -1391C, -566$(GT)_{20}$ and -386C (see Table 3.4, section 3.2.18). Hence, pgl3b-Brn3c-3.6 carries the variant allele, C, at -1391 and was also known as construct p-Luc-1391C. In the first instance, construct p-Luc-1391C (pgl3b-Brn-3c-3.6) was used as a template in a site-directed mutagensis (SDM) reaction (QuikChange®, Stratagene)

(see section 3.2.2.2) with primers -1391aSDM-S: 5' GCAGCGTAGTCGAGGTCCAGGATTC 3' and -1391tSDM-AS: 5' GAATCCTGGACCTCGACTACGCTGC 3' to modify the variant allele C at position -1391 to the wild type allele, A, to form construct p-Luc-1391A. Thus, generating two different Brn-3c luciferase-reporter gene constructs that differed in the nature of the -1391 allele (see Table 3.4, section 3.2.18). For convenience construct p-Luc-1391A was also referred to as p-Luc-SNPG1$_{-566GT20}$ in some experiments.

Construct p-Luc-SNPG1$_{-566GT20}$ (p-Luc-1391A) was used as a template to form constructs p-Luc-SNPG1$_{-566GT16}$ and p-Luc-SNPG1$_{-566GT23}$ that both carried allele SNPG1 at position -3432 together with the native 5' haplotype at -3457 and -3495 [$(GA)_2$ and $(GT)_{15}$, respectively] but differed in the number of dinucleotide repeats at position -566. Briefly, primers 3cP4-S and 3cP10-AS (for sequence see Table 3.1, section 3.2.18) were used in a PCR reaction to generate a 710bp fragment spanning the -566$(GT)_n$ repeat using genomic DNA samples of genotype that had been determined from the initial SSCP screen (see section 3.2.2.3 for details of the PCR reaction). Post-PCR the resultant amplicons were sub-cloned into pGEM®-T easy vector, sequenced to confirm genotype: [-566$(GT)_{16}$: -386C] and [-566$(GT)_{23}$: -386C], cleaved with *SacII*, and the resulting 380bp restriction fragments ligated into Sac*II* digested p-Luc-SNPG1$_{-566GT20}$. A PCR screen with primers 3cP5-S and 3cP6-AS confirmed the correct (forward) orientation of the *SacII* inserts in these constructs (see Table 3.4, section 3.2.18).

An additional set of Brn-3c promoter-luciferase reporter gene constructs were devised that carried allele SNPG2 at position -3432 together with the native 5' haplotype at -3457 and -3495 [$(GA)_2$ and $(GT)_{21}$, respectively] that also differed in the number of dinucleotide repeats at position -566. Briefly, a pGEM®-T easy clone carrying a 398bp genomic DNA insert with allele SNPG2 and native 5' haplotype defined as -3495$(GT)_{21}$: -3457$(GA)_2$: -3432GGC$(G)_9$ (see section 3.2.2.4 for generation of 398bp fragment) was digested with *SacI* and *SpeI* and the resultant 342bp fragment ligated into *SacI* and *SpeI* cut p-Luc-SNPG1$_{-566GT16}$ and p-Luc-SNPG1$_{-566GT23}$ to form constructs p-Luc-SNPG2$_{-566GT16}$ and p-Luc-SNPG2$_{-566GT23}$, respectively (see Table 3.4, section 3.2.18). All constructs were sequenced to verify authenticity.

3.2.8.2 Expression constructs.

The mammalian expression vector pEVR2-Sp1 that contains the complete human cDNA for SP1 has been described previously (Hagen et al, 1992) and was kindly

provided by Professor Guntram Suske, University of Marburg, Germany. The pEVR2-Sp1 plasmid was digested with *PvuI* and then *XbaI* to generate a SP1 cDNA insert defined by *XbaI* overhangs that could be clearly isolated from the pEVR2 backbone. This was ligated into *XbaI* cut pSi (Promega) to form the mammalian expression construct pSi-SP1. The correct (forward) orientation of the SP1 cDNA insert in pSi-SP1 was confirmed by *BglII* digest (inserts in the forward orientation give two bands of 4.6Kb and 3.0Kb).

The *Drosophilia* expression vector pPac-USp3 containing the complete human SP3 cDNA has been described previously (Dennig et al, 1996) and was kindly provided by Professor G.Suske as before. The SP3 cDNA insert was digested from pPac-USp3 by *SpeI* / *XhoI* cleavage. The pSi mammalian expression vector (Promega) was cleaved with *XbaI* / *SalI* to generate compatible 5' overhangs for acceptance of the *SpeI* / *XhoI* SP3 cDNA insert upon ligation. This formed the mammalian expression construct pSi-SP3. The correct (forward) orientation of the SP3 cDNA insert in pSi-SP3 was confirmed by *BglII* digest (inserts in the forward orientation give two bands of 4.4Kb and 1.6Kb).

The mammalian dominant negative human SP1 expression vector pEBGN-SP1 encoding a fusion protein of glutathione S-transferase (GST) linked to the DNA binding domain (DBD) of SP1 has been described previously (Petersohn and Thiel, 1996; Lietz et al, 2003) and was kindly provided by Professor Gerald Thiel, University of the Saarland Medical Centre, Germany. Using plasmid pEBGN-SP1 as a template, primer Sp1DomNeg-S: 5' **ATG**CTGGTTCCGCGTGGATCAC 3' designed to include a 5'ATG for an in-frame methionine and primer Sp1DomNeg-AS: 5 ' CCCTCACTCTAGAGTCGC 3' were used to amplify a fragment containing the cDNA for the SV40 nuclear localisation signal (NLS) linked to the DBD of SP1. PCR was performed using the BD Advantage ™ 2 PCR Kit (BD Biosciences) for high fidelity PCR (see section 3.2.2.5). The resultant fragment was gel purified (QIAquick gel extraction kit, QIAGEN), sub-cloned into the pGEM®-T easy vector (Promega) and sequenced to ensure correct. The SV40 NLS - SP1-DBD cDNA chimera was then cleaved from pGEM®-T easy vector by *EcoRI* digest and ligated into *EcoRI* cut pSi (Promega); this formed the mammalian dominant negative expression vector for human SP1 termed, pSi-DBD-SP1. The correct (forward) orientation of the DBD-SP1 cDNA insert in pSi-DBD-SP1 was confirmed by *BglII* digest (inserts in the forward orientation give two bands of 3.0Kb and 1.3Kb).

3.2.8.3 pGL3-promoter constructs.

Oligonucleotides with 5' overhangs (either *Bgl II* or *Nhe I*) spanning: common sequence variants in the Brn-3c promoter; containing a consensus transcription factor binding site; or a mutated transcription factor binding site, were 5'-end phosphorylated in a standard T4 polynucleotide kinase reaction (see section 3.2.9.5). Phosphorylated oligonucleotides were purified by ethanol precipitation (Sambrook et al, 1989), diluted in ddH_2O and ligated into *Bgl II* or *Nhe I* digested pGL3-promoter vector (Promega) as appropriate. Insertion of the required oligonucleotide into pGL3-promoter vector was confirmed by replica-plating of recombinant clones and hybridisation with [γ-^{32} P] labelled single stranded oliogonucleotide as probe. Recombinant clones confirmed with oligonucleotide insert were screened for copy number of insert (oligonucleotides with 5'-overhangs will ligate to each other in addition to the linearised vector during the ligation reaction) by double restriction enzyme digest with either *Nhe I* and *Nco I* or *Sac I* and *Nco I* as appropriate (see Table 3.5, section 3.2.18).

3.2.9 DNA manipulations.

3.2.9.1 TA-cloning.

Post-PCR amplified fragments were directly cloned into the pGEM®-T easy vector (Promega) for sequencing. The pGEM®-T easy vector is linearised at the multiple cloning region with single 3'-thymidine overhangs that facilitates the insertion of PCR products with adenosine appendages at their 3'-ends. DNA polymerases such as *Taq* have terminal transferase activity and characteristically append the 3'-ends of amplified fragments in a manner that is independent of template. Typically, the PCR reaction products were chloroform extracted for purification (Sambrook et al, 1989) and purified amplicons sub-cloned into the pGEM®-T easy vector using the 2x rapid ligation buffer kit (Promega). Briefly, a 2µl aliquot of the purified PCR product was mixed with 5µl rapid ligation buffer, 25ng pGEM®-T easy vector and 1µl T_4 DNA ligase in a reaction volume of 10µl. Contents were incubated at room temperature for 2-3 hours or overnight at 4°C.

3.2.9.2 Restriction endonuclease digestion.

Restriction endonuclease digest was generally used to prepare staggered or blunt ends of insert and vector plasmid DNA for sub-cloning or to screen for insert orientation in

recombinant clones. Typically 1-3μg of plasmid DNA was digested with the appropriate restriction enzyme (maximum amount 10% v/v) in 1x of the appropriate restriction enzyme buffer (Promega) in a total volume of 20-30μl. For double restriction enzyme digests if the two restriction enzymes were not compatible in the same restriction enzyme buffer or the multicore buffer (Promega), the digests were performed sequentially. Ethanol precipitation (SamBrook et al, 1989) was used to precipitate the DNA after the first digest and the precipitated DNA was re-suspended in ddH_2O for the second digest. Reactions were left for 1-2 hours or overnight at 37°C (or at the required temperature as appropriate) and analysed by agarose gel electrophoresis (see section 3.2.12.1).

3.2.9.3 Phosphatasing of plasmid vectors.

Phosphatase treatment was used to facilitate vector DNA for acceptance of insert DNA prior to ligation. After restriction endonuclease digest vector DNA was de-phosphorylated with calf intestinal alkaline phosphatase (AP) to help prevent re-ligation of digested vector DNA; AP catalyses the hydrolysis of the 5'-phosphate groups from DNA. Typically, 1μl of AP was added direct to the digested vector DNA (ensuring that the total enzyme content did not exceed 10 % v/v) with 1x AP buffer (Promega) in a total volume of 25μl and incubated at 55°C for 45 minutes. The AP was then deactivated by heating to 65°C for 15 minutes.

3.2.9.4 Ligation of DNA.

Digested, purified* vector and insert DNA were ligated using a molar ratio of either 1:3 or 1:7 DNA molecules, respectively with 1x T_4 DNA ligase buffer (Promega) and 3U of T_4 DNA ligase in a total volume of 10μl. Ligations were incubated at 16°C overnight.

*Vector and insert DNA were purified by recovering from low melting point agarose gels, see section 3.2.11.14.

3.2.9.5 Phosphorylation of oligonucleotides.

5'-end phosphorylation was used to prepare oligonucleotides for sub-cloning. Annealed oligonucleotides were 5'end phosphorylated in a standard T4 polynucleotide kinase reaction with dATP. Briefly, 5μg of annealed oligonucleotide was combined with 2ul of 10mM dATP, 5U of T4 polynucleotide kinase and 1x standard kinase buffer (Promega) in a total volume of 20μl and left to incubate at 37°C for 30 minutes.

3.2.9.6 Bacterial transformation.

Competent *Escherichia Coli (E. Coli)* XL1-Blue cells were prepared according the method of Hanahan, 1985 with minor modifications. For transformation, 5µl of the ligation mix was pipetted into 50µl of freshly thawed competent cells and incubated on ice for 45 minutes. Cells were subject to a heat-shock by incubating for 2 minutes at 42°C, 200µl of LB medium was added followed by incubation at 37°C for 30 minutes in a rotating shaker at 200rpm. Cells were plated onto standard LB agar-AMP plates containing 50µg/ml ampicillin or if recombinant clones were to be selected by blue-white colour screening, LB agar-AMP plates containing 50µg/ml ampicillin supplemented with 0.02mM IPTG, and 20µg/ml X-gal. Plates were incubated at 37°C overnight.

3.2.10 Screening of recombinant clones.

Recombinant clones were screened by blue-white colour screening, restriction endonuclease digest (see section 3.2.9.2), by PCR reaction or by hybridisation as appropriate.

3.2.10.1 By blue-white colour screening coupled with PCR.

Recombinant clones were identified on indicator plates by blue-white colour screening. The cloning of an insert into the pGEM®-T easy vector (Promega) results in insertional inactivation of the α-peptide coding region of the enzyme β-galactosidase. Consequently recombinant clones containing the required insert give rise to white (or pale blue) colonies on indicator plates. Routinely several white (or pale blue) colonies were picked and streaked out onto a master plate (typically a standard LB-agar AMP plate) and screened for plasmids containing the relevant insert by PCR direct from the bacterial colonies using primers that were used in the original PCR amplification for that insert. Positive clones were grown up from the master plate at 37°C in 5ml of LB medium containing 50µg/ml ampicillin, in a rotating shaker at 200rpm, overnight.

3.2.10.2 By PCR-screening.

To confirm insertion of the required insert and / or insert orientation in pGL3-Basic vector (Promega) recombinant clones were screened by PCR reaction. Routinely several white colonies were picked and streaked out onto a master plate (a standard LB-agar

AMP plate) and screened for plasmids containing the relevant insert by PCR direct from the bacterial colonies using primers as appropriate (see Table 3.4, section 3.2.18). Positive clones were grown up from the master plate at 37°C as before.

3.2.10.3 By hybridisation.

To confirm insertion of the required oligonucleotide into pGL3-promoter vector recombinant clones were screened by replica-plating and hybridisation with [γ-^{32} P] labelled single stranded oliogonucleotide as probe. Routinely, recombinant clones were replica-plated to a Hybond™-N+ nylon membrane (Amersham Biosciences) on a standard LB-agar AMP plate and incubated at 37°C overnight. The membrane was removed from the LB-AMP plate and placed on top of two sheets of Whatman 3MM paper pre-wetted with 10% SDS and left for 3 minutes, transferred to Whatman 3MM paper pre-wetted with 1x denaturing solution and left for 7 minutes, quickly rinsed in 3x SSC and DNA was cross-linked to the membrane with an UV stratalinker. The membrane was air-dried and stored in Saran wrap at 4°C until ready to hybridise.

The hybridisation solution was freshly prepared and pre-warmed to the required hybridisation temperature. The membrane was carefully rolled and placed inside a hybridisation bottle and pre-wetted with 3x SCC. The SCC was discarded and the membrane covered with pre-warmed hybridisation solution (5-10mls for a small hybridisation bottle) and set on a carousel in a hybridisation oven at the required hybridisation temperature. [γ-^{32} P] labelled double stranded oliogonucleotide (50ng) was denatured by heating to 95°C for 3 minutes and placed on ice ready for use as single stranded probe. Probe was pippetted into the hybridisation bottle and left to hybridise overnight. Post-hybridisation, probe was discarded and membrane rinsed with pre-warmed wash solution (3x SSC, 0.1% SDS). In order to remove non-specific binding of probe the membrane was incubated with a new aliquot of pre-warmed wash solution for 15 minutes, the wash solution was discarded and this step repeated. The membrane was removed from the hybridisation bottle and excess liquid removed with a paper towel without allowing the membrane to dry out. The membrane was wrapped in Saran wrap and insertion of required oligonucleotides into pGL3-promoter vector confirmed by autoradiography of the membrane with Hyperfilm™ MP film (Amersham Biosciences) at -80°C. Recombinant clones containing the empty pGL3-promoter vector were used as a negative control to confirm binding of [γ-^{32} P] labelled probe was due to

recombinant clones containing the required insert cross-linked to the membrane. Positive clones were grown up from the master plate at 37°C as before.

3.2.11 DNA preparation.

3.2.11.1 Genomic DNA preparation from human blood.

Blood was thawed at room temperature. 3mls of thawed blood was diluted to 25mls in a 50ml Falcon® tube with cold Sucrose-Tris lysis buffer (0.32M Sucrose, 10mM Tris-HCl pH 7.5, 5mM $MgCl_2$ and 1% Triton X-100), inverted to mix and spun at 2500rpm at 4°C for 10 minutes. The supernatant was discarded and traces of lysis buffer removed from the tube with a paper tissue. The pellet was re-suspended in 68mM NaCl, 22mM EDTA pH 8.0, 0.5% SDS with 1mg proteinase K in a total volume of 5mls, transferred to a 12ml Falcon® tube and incubated at 37°C overnight. After which, 5mls of equilibrated phenol was added, the tube inverted gently several times and spun at 2000rpm for 5 minutes. The upper aqueous phase was carefully removed to a clean tube and mixed with 5mls of phenol:chloroform:isoamyl-alcohol (25:24:1; ReadyRed, Qbiogene) and spun as before. The aqueous phase was removed to a new tube, 5mls chloroform:isoamyl-alcohol (24:1; ReadyRed, Qbiogene) added and the tube was mixed and spun as before. The aqueous phase was removed to a new tube, 0.5ml of 3M sodium acetate added, the tube was swirled to mixed and 12.5ml of cold ethanol quickly added. The tube was gently inverted several times until DNA precipitated. Precipitated DNA was transferred to an eppendorf tube, centrifuged at 13,000rpm for 1 minute and excess liquid removed. The pellet was air dried for 10-15 minutes and re-suspended in 200µl ddH_20 on a rotator for 2 days at 4°C. DNA concentration was determined by spectroscopy at 260nm and DNA was stored at -20°C with concentrated DNA stocks stored at -80°C.

3.2.11.2 Small-scale preparation of plasmid DNA from transformed *E.Coli.*

Plasmid DNA was purified from 3-5ml of overnight bacterial culture using the Nucleon® plasmid mini-prep kit (Nucleon Biosciences) or the QIAprep® spin mini-prep kit (QIAGEN) according to the manufacturer's instructions. DNA was quantified by spectrophotometry at 260nm and running an aliquot on an agarose gel. If required a glycerol stock was prepared of the liquid culture by vortexing 1ml of culture with 500µl of autoclaved glycerol and storing at -80°C.

3.2.11.3 Medium-scale preparation of plasmid DNA from transformed *E.Coli.*

Plasmid DNA was purified from 50ml of overnight bacterial culture using the QIAGEN® plasmid midi purification kit according to the manufacturer's instructions with minor modifications. DNA was quantified by spectroscopy at 260nm and running an aliquot on an agarose gel. If required a glycerol stock was prepared as described previously (3.2.11.2). Plasmid integrity was confirmed by restriction endonuclease digest.

3.2.11.4 Purification of DNA from low melting point agarose gels.

Prior to ligation a specific PCR product, digested plasmid insert or vector DNA was recovered from a standard low melting point agarose gel. 0.7[1]-2.4% low melting point agarose gels were prepared with 1x TAE buffer containing 0.25µg/ml ethidium bromide. DNA to be recovered was loaded with 1x DNA loading buffer and electrophoresed alongside 3-5µl of 1KB DNA ladder (Invitrogen) as appropriate. DNA was briefly visualised with a 3UV™ transilluminator GelDoc-It imaging system (Jencons-PLS) and the required fragments excised from the gel and purified using the QIAquick gel extraction kit (QIAGEN) with a microcentrifuge according to the manufacturer's instructions.

[1]0.7% low melting point agarose gels were prepared and electrophoresed at 4°C.

3.2.12 Electrophoresis.

3.2.12.1 Agarose gel electrophoresis of DNA.

0.7-3% agarose gels were prepared with 1x TAE buffer containing 0.25µg/ml ethidium bromide. Samples were loaded with 1x DNA loading buffer and electrophoresed alongside 3-5µl of 1KB DNA ladder (Invitrogen) as appropriate. DNA was visualised at 302nm with a 3UV™ transilluminator GelDoc-It imaging system (Jencons-PLS).

3.2.12.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using the Jencons-PLS vertical electrophoresis unit (Model V10-CDC). Resolving and stacking gels were prepared with the Jencons-PLS gel-casting unit according to the manufacturer's instructions. OC-2 nuclear extract (24µg) and 10µl Rainbow™ coloured protein molecular weight markers (Amersham Biosciences) were made up to 20 µl with 1x SDS gel-loading buffer, heated at 95°C for 2 minutes, immediately loaded onto a 5%

stacking gel and electrophorised on an 8% resolving gel at 150V in 1x western running buffer until the appropriate pre-stained molecular weight markers had separated within the gel. The gel was immunoblotted as described in section 3.2.13.

3.2.13 Western Immunobloting.

Following SDS-PAGE, the protein was transferred from the polyacrylamide gel to a Hybond™-C nitrocellulose membrane using a Bio-RAD Trans-Blot® electrophoretic transfer cell. The cassette containing the gel and nitrocellulose membrane was assembled with 3MM Whatman paper and submerged in blotting buffer according to the manufacturer's guidelines. Transfer was performed at a constant current of 210mA for 3 hours; successful transfer of proteins was confirmed by visualisation of the Rainbow™ coloured protein molecular weight markers on the membrane. Post-transfer the membrane was blocked for 1 hour at room temperature in TBS-T containing 10% non-fat milk powder (Marvel), washed for 5 minutes in TBS-T and subsequently incubated for 1 hour at room temperature with primary antibody, mouse monoclonal anti-SP1 antibody [SP1 (1C6) X, Santa Cruz Biotechnology Inc.], diluted 1:10,000 fold in TBS-T containing 5% non-fat milk powder. After which, the membrane was rinsed twice in TBS-T followed by two 5 minute incubations in TBS-T containing 5% non-fat milk powder and then incubated for 1 hour at room temperature with secondary antibody, horseradish peroxidase (HRP)-conjugated goat anti-mouse antibody (DAKO) diluted 1:1000 fold in TBS-T containing 5% non-fat milk powder. The membrane was then rinsed in TBS-T for 15 minutes followed by two 5 minute washes. All procedures were performed on a mixer platform with gentle agitation. Protein bound with anti-SP1 antibody on the membrane was identified by detecting HRP activity with a chemiluminescence system (ECL Plus Western blotting detection reagents, Amersham Biosciences) according to the manufacturer's instructions and visualised by exposing the membrane to autoradiography film, Hyperfilm™ ECL™ (Amersham Biosciences).

3.2.14 Transient transfection and luciferase assay.

The day before transfection, OC-2 cells were plated into 6-well plates at 2 x 10^5 cells per well. Cells were transfected according to the calcium phosphate precipitation method (Gorman, 1985). Typically, 0.5μg of *firefly* luciferase reporter plasmid was transfected with 10 or 100ng *Renilla* expression plasmid (pRL-SV40 or phRL-null as

appropriate) as an internal control to normalise for differences in transfection efficiency. In co-transfections 0.5μg - 3μg of expression plasmid (either pSi-SP1, pSi-SP3, pSi-DBS-SP1 or CMV5-Gfi-1) was titrated in as appropriate and the empty expression plasmid (either pSi or pCi as appropriate) was used as a negative control to ensure that regulation was not due to the plasmid backbone and to standardise DNA concentrations between wells. Additional experiments were performed with the empty *firefly* luciferase reporter plasmid to ensure that regulation was specific to the Brn-3c promoter insert. Twenty-four hours after transfection the medium was removed and a standard glycerol shock was performed (15% glycerol). Cells were harvested 24 hrs after glycerol shock and luciferase activities determined according to the Dual-Luciferase® Reporter Assay System (Promega). Luminescence was measured using a Turner TD-20e luminometer. Each experiment was performed in triplicate and repeated at least twice with two different preparations of DNA. Error bars on charts represent the standard error of the mean.

3.2.15 Genotyping -3432poly-G polymorphism.

The -3432poly-G-polymorphsim was genotyped by direct sequencing in combination with PAGE analysis of radioactive PCR products to size poly-G alleles both procedures were optimised to produce the protocol described below.

3.2.15.1 Direct sequencing.

Primers 3cP11-S (5'-GCATAAATGCCCACATAGTCC-3') and 3cP12-AS (5'-GCAATTTGTAACAGCCTCTATGC-3') were used to amplify a 285bp fragment* containing the -3432poly-G polymorphism. PCR was carried out in 96-well format using 100ng genomic DNA per reaction. Each PCR reaction comprised 0.2μM of each primer, 200μM dNTPs, 1x GeneAmp PCR Gold buffer (Applied Biosystems), 2.0mM $MgCl_2$, and 0.625U of AmpliTaq Gold (Applied Biosytems) in a total volume of 25μl. Thermal cycling was performed with an Eppendorf Mastercycler Gradient using an initial denaturation step of 95°C for 10 minutes followed by denaturation at 94°C for 30 seconds, annealing at 59°C for 30 seconds, extension at 72 °C for 30 seconds, for 30 cycles, with a final extension of 70°C for 10 minutes. Post-PCR, PCR products were purified using the QIAquick PCR purification kit (QIAGEN) with a microcentrifuge according to the manufacturer's instructions and 5μl of the purified PCR mix was

electrophoresed on a 1.5% agarose gel to check successful amplification of the required band. Purified samples were diluted in ddH_2O as appropriate, and 100ng of each purified sample was sequenced on an ABI 3730 DNA analyser with BigDye terminator chemistries version 3.1, MRC geneservice (Babraham Bioincubator, Cambridge, U.K.) using primer 3cP12-AS in the sequencing reaction.

3.2.15.2 PAGE analysis of radioactive PCR products.

Length of poly-G alleles was determined by incorporating a [α^{32}P]-dCTP radiolabel in the amplification of an 181bp product* spanning the -3432poly-G polymorphism using primers G3cP13-S (5'-GATTGTAATTTAATGCCATGGTG-3') and 3cP12-AS (5'-GCAATTTGTAACAGCCTCTATGC-3') in the PCR reaction. Each PCR reaction comprised 100ng genomic DNA, 0.2µl [α^{32}P]-dCTP, 0.2µM of each primer, 200µM dATP, 200µM dGTP, 200µM dTTP and 100µM dCTP with 1x GeneAmp PCR Gold buffer (Applied Biosystems), 2.0mM $MgCl_2$, and 0.625U of AmpliTaq Gold (Applied Biosytems) in a total volume of 25µl. Thermal cycling was performed with an Eppendorf Mastercycler Gradient as described above (3.2.15.1) but with an annealing temperature of 56°C. Post-PCR, 5µl of the PCR reaction was mixed with 5µl formamide gel-loading buffer, heated to 95 °C for 5 minutes and immediately placed on ice prior to loading on a 6% denaturing polyacrylamide gel (SequaGel 6, National Diagnostics) in 1x TBE alongside plasmid markers** of known length. Samples were electrophorised using a Life technologies Inc. vertical electrophoresis unit (Model S2) at 60Watts for 2 hours 30 minutes and bands were visualised by autoradiography of the dried gel on BioMax MR film (Kodak) at -80°C.

*based on containing poly-G allele $(G)_{11}$ which is of equivalent length to alleles SNPG4 and SNPG5 (see Reference sequence: Ensembl transcript ID ENST00000230732; Appendix B).

**Plasmid markers were produced by using 100ng plasmid DNA in the PCR reaction. Plasmids were pGEM®-T easy constructs carrying specific -3432poly-G alleles that had been confirmed by sequencing.

3.2.16 Genotyping -1391A>C.

The -1391A>C single nucleotide substitution was genotyped by PCR-*TaqI* restriction digest. Initially the PCR reaction was optimised for production of a single, intense band of the required size by varying the following parameters: the magnesium concentration and annealing temperature of the reaction, the DNA template concentration, the number of PCR cycles, the length of the annealing and extension stage of the PCR reaction and

the type of DNA polymerase used [*Taq* DNA polymerase (Promega) versus AmpliTaq Gold (Applied Biosystems). The optimised protocol consisted of a 25µl PCR reaction per sample with 100ng genomic DNA. PCR was carried out in 96-well format using 0.2µM each primer (3cP1-S and 3cP3-AS; for sequence see Table 3.1, section 3.2.18), 200µM dNTPs, 1x GeneAmp PCR Gold buffer (Applied Biosystems), 2.0mM $MgCl_2$, and 0.625U of AmpliTaq Gold (Applied Biosystems). Thermal cycling was performed as described previously (3.2.15.1) but with an annealing and an extension stage of 45 seconds. Specificity of the PCR reaction was confirmed by electrophoresing a 5µl aliquot of a few samples on a 1.5% agarose gel and visualisation of a single 596bp band. Post-PCR the *TaqI* digest was performed by adding an aliquot of *TaqI* directly to the PCR reaction mix at a final concentration of 4% and incubating at 65°C for 2½ hours. Restriction fragments were then visualised on a 2% agarose gel stained with ethidium bromide (0.25µg/ml) and genotypes allocated according to the bands present: homozygous wild-type (AA) 81, 123 and 392bp, homozygous variant (CC) 204 and 392bp and heterozygous 81, 123, 204 and 392bp.

3.2.17 Statistics.

Paired or unpaired *t*-tests, as appropriate, were performed to compare data sets obtained from transient transfection assay; a p-value <0.05 was considered significant. Analysis was performed using SPSS 12.0.1 statistical software package.

The Chi-square (χ^2) test was used to compare genotype and allelic frequencies between patient and general population sample groups for the preliminary association study; a p-value <0.05 was considered significant (for χ^2 probability table see Swinscow and Campbell, 2002). χ^2 analysis was performed using the formula:

$$\chi^2 = (O - E)^2 / E$$

Where **O** is the observed frequency distribution and **E** is the expected frequency distribution (see Swinscow and Campbell, 2002)

Genotype distribution assuming Hardy-Weinberg equilibrium was calculated using the Hardy-Weinberg equation:

$$\mathbf{P^2\,(AA) + 2pq\,(Aa) + q^2\,(aa) = 1}$$

Where **A** and **a** are two alleles of gene; **A** is the dominant allele in the population and **a** is the recessive allele.

The frequency of the **A** allele = **p** and the frequency of the **a** allele = **q**

Where $\mathbf{p + q = 1}$.

Statistical levels of significance for deviation from expected Hardy-Weinberg equilibrium were tested by χ^2 analysis as described previously and a p-value <0.05 was considered significant.

3.2.18 Tables.

Table 3.1: Summary of primers for PCR-SSCP analysis of the Brn-3c gene.

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3cP1-S	Promoter	5' CTCTCAGCGGAGGCAGTGG 3'	1	216	62.3	-
3cP1-AS		5' TGAGGGCTCTCGAGGTTAGC 3'				
3cP2-S	Promoter	5' GCAGTGCTAACCTCGAGAGC 3'	2	210	62.3	-
3cP2-AS		5' CCACCGTATGGAGTAGGTGATGT 3'				
3cP3-S	Promoter	5' AACATCACCTACTCCATACGGTG 3'	3	220	62.3	-
3cP3-AS		5' CCGTCTAAGGAAGCTTGTGGAG 3'				
3cP4-S	Promoter	5' CTCCACAAGCTTCCTTAGACG 3'	4	276	62.3	-
3cP4-AS		5' CCAAGCTGTCCACCAGAGTC 3'				
3cP5-S	Promoter	5' GGACTCTGGTGGACAGCTTG 3'	5	209	63.4	-
3cP5-AS		5' TGTCCCAGCTCGAACTGCC 3'				
3cP6-S	Promoter	5' GGCAGTTCGAGCTGGGACA 3'	6	256	62.3	+
3cP10-AS		5' TCACGCGCCGGGGACTCG 3'				
3cP9-S	Promoter	5' AAGGCGCGCCGCTAGCTG 3'	7	295	54.2	+
3cP7-AS		5' GAAAGGCTGCTTGGAGTTCAT 3'				
3cP7-S	Promoter	5' CTTCCTGGCAGGCTGCTTGT 3'	8	205	62.3	-
3cP7-AS		5' GAAAGGCTGCTTGGAGTTCAT 3'				
3cP11-S	Promoter*	5' GCATAAATGCCCACATAGTCC 3'	15	130	57.5	-
3cP11-AS		5' CCACCACCATGGCATTAAA 3'				
Ex1-S	Exon 1	5' AAGCCTGATTCCATGTCACC 3'	9	228	58	-
Ex1-AS		5' CATGAAGCTAGTGCCTGTCAA 3'				
3cI-S	Intron	5' GAGCATAATTACCGCTCTAAGGC 3'	10	287	62.3	-
3cI-AS		5' ACAGCACGGTCAGTAGGGAATAC 3'				
Ex2F1-S	Exon 2	5' TTCCCTACTGACCGTGCTGT 3'	11	302	58	-

Ex2F1-AS		5' ACTCAGCGTGGGCGAGATGT 3'				
Ex2F2-S	Exon 2	5' TTCGTCCACCGTGCCCATCT 3'	12	271	63	-
Ex2F2-AS		5' CTCCACGTCGCTGAGGCATG 3'				
Ex2F3-S	Exon 2	5' TCAGCGACGTGGAGTCAGA 3'	13	254	63	-
Ex2F3-AS		5' CTCTCGGTAGGCGGCCTCG 3'				
Ex2F4-S	Exon 2	5' CGAGGCCGCCTACCGAGAG 3'	14	261	63	-
Ex2F4-AS		5' TGCCGCAATCAGTGGACAG 3'				

*Fragment 15 spans -3495$(GT)_n$ within the Brn-3c promoter. All other promoter fragments are within 1.5Kb based on the first A of the ATG translation start for Brn-3c as +1.

Table 3.2: Summary of oligonucleotides used in EMSA analysis.

Oligonucleotide name	Type: allele oligonucleotide spans / consensus sequence	Sequence (5' overhangs are denoted in red).	Length (bp)
(G)9-S	-3432 $(G)_9$ *	5' GATCTGGTGGTGGGGGGGGGTGGGTAAGT 3'	29
(G)9-AS	-3432 $(G)_9$ *	5' GATCACTTACCCACCCCCCCCCACCACCA 3'	29
(G)10-S	-3432 $(G)_{10}$	5' GATCTGGTGGTGGGGGGGGGGTGGGTAAGT 3'	30
(G)10-AS	-3432 $(G)_{10}$	5' GATCACTTACCCACCCCCCCCCCACCACCA 3'	30
(G)11-S	-3432 $(G)_{11}$	5' GATCTGGTGGTGGGGGGGGGGGTGGGTAAGT 3'	31
(G)11-AS	-3432 $(G)_{11}$	5' GATCACTTACCCACCCCCCCCCCCACCACCA	31
(G)12-S	-3432 $(G)_{12}$	5' GATCTGGTGGTGGGGGGGGGGGGTGGGTAAGT 3'	32
(G)12-AS	-3432 $(G)_{12}$	5' GATCACTTACCCACCCCCCCCCCCCACCACCA 3'	32
SNPG1-S	-3432 SNPG1	5' GATCTGGTGGTGGGGGGGGCGTGGGTAAGT 3'	30
SNPG1-AS	-3432 SNPG1	5' GATCACTTACCCACGCCCCCCCCCACCACCA 3'	30
SNPG2-S	-3432 SNPG2	5' GATCTGGTGGTGGCGGGGGGGGGGTGGGTAAGT 3'	32
SNPG2-AS	-3432 SNPG2	5' GATCACTTACCCACCCCCCCCCGCCACCACCA 3'	32
SNPG4-S	-3432 SNPG4	5' GATCTGGTGGTGGTGGGGGGGGGTGGGTAAGT 3'	31
SNPG4-AS	-3432 SNPG4	5' GATCACTTACCCACCCCCCCCCACCACCACCA 3'	31
SNPG5-S	-3432 SNPG5	5' GATCTGGTGGTGGGGCGGGGGGTGGGTAAGT 3'	31
SNPG5-AS	-3432 SNPG5	5' GATCACTTACCCACCCCCCGCCCCACCACCA 3'	31
1391C-S	-1391 C	5' GATCAGCGTAGTCGCGGTCCAGGATTCC 3'	28
1391C-AS	-1391 C	5' GATCGGAATCCTGGACCGCGACTACGCT 3'	28
1391A-S	-1391 A	5' GATCAGCGTAGTCGAGGTCCAGGATTCC 3'	28
1391A-AS	-1391 A	5' GATCGGAATCCTGGACCTCGACTACGCT 3'	29

CPBP-S	CPBP consensus site[1]	5' GATCTGACCCCACCCATGAGCCTGAGAAGTGC 3'	32
CPBP-AS	CPBP consensus site[1]	5' GATCGCACTTCTCAGGCTCATGGGTGGGGTCA 3'	32
SP1-S	SP1 consensus site	5' GATCATTCGATCGGGGCGGGGCGAGC 3'	26
SP1-AS	SP1 consensus site	5' GATCGCTCGCCCCGCCCCGATCGAAT 3'	26
SP1 Mut-S	Mutated SP1 binding site	5' GATCATTCGATCGG**TT**CGGGGCGAGC 3'	26
SP1 Mut-AS	Mutated SP1 binding site	5' GATCGCTCGCCCCG**AA**CCGATCGAAT 3'	26
Egr-S	Consensus Egr binding site	5' CTAGGGATCCAGCGGGGGCGAGCGGGGGCGA 3'	30
Egr-AS	Consensus Egr binding site	5' CTAGTCGCCCCCGCTCGCCCCCGCTGGATCC 3'	30
Egr Mut-S	Mutated Egr binding site	5' CTAGGGATCCAGC**TA**GGGCGAGC**TA**GGGCGA 3'	30
Egr Mut-AS	Mutated Egr binding site	5' CTAGTCGCCC**TA**GCTCGCCC**TA**GCTGGATCC 3'	30
Gfi-1-S	Consensus Gfi-1 binding site	5' CTAGACCATCACCACATAAATCACTGCCTATCCTGTG 3'	37
Gfi-1-AS	Consensus Gfi-1 binding site	5' CTAGCACAGGATAGGCAGTGATTTATGTGGTGATGGT 3'	37
Gfi-1 Mut-S	Mutated Gfi-1 binding site	5' CTAGACCATCACCACATAA**C**TCACTGCCTATCCTGTG 3'	37
Gfi-1 Mut-AS	Mutated Gfi-1 binding site	5' CTAGCACAGGATAGGCAGTGA**G**TTATGTGGTGATGGT 3'	37
Bar-Ex1-S	Section of exon 1 of Barhl1	5' AGCGACTGCTCTTCGCCAG 3'	19
Bar-Ex1-AS	Section of exon 1 of Barhl1	5' CTGGCGAAGAGCAGTCGC 3'	18

* $(G)_9$ was later determined to be a sequencing error and not a true allele. GATC: *BglII* 5'overhang; CTAG: *NheI* 5' overhang. Alleles are highlighted in black and blue bold. Where mutated binding sites are used, mutated nucleotides are highlighted in black bold. [1]Koritschomer et al, 1997.

Table 3.3: Summary of primers used in sequencing analysis.

Primer	Sequence	Reason for use
Rvprimer3*	5' CTAGCAAAATAGGCTGTCCC 3'	Confirm -3495$(GT)_n$, -3457$(GA)_n$ and -3432poly-G polymorphism genotype in Brn-3c promoter inserts.
GLprimer2*	5' CTTTATGTTTTTGGCGTCTTCCA 3'	Confirm -566$(GT)_n$ and -386C>A genotypes in Brn-3c promoter inserts. Also provides sequence of GC box and TATA box in Brn-3c promoter inserts.
3cP4-S	5' CTCCACAAGCTTCCTTAGACG 3'	
3cP5-S	5' GGACTCTGGTGGACAGCTTG 3'	
3cP2-AS	5' CCACCGTATGGAGTAGGTGATGT 3'	Confirm -1391A>C genotype in Brn-3c promoter inserts.
3cP1-S	5' CTCTCAGCGGAGGCAGTGG 3'	
SP6**	5' TATTTAGGTGACACTATAG 3'	Confirm sequence in pGEM®-T easy inserts.
T7 forward**	5' TAATACGACTCACTATAGGG 3'	Confirm sequence in pGEM®-T easy inserts.

* Purchased from Promega.

** Provided by the sequencing service (for details see section 3.2.7).

Table 3.4: Summary of human Brn-3c promoter constructs

Construct	-3495(GT)$_n$	-3457(GA)$_n$	-3432poly-G polymorphism	-1391A>C	-566(GT)$_n$	-386C>A	Construct origin.	[redacted] clones.	Primers used in sequencing analysis to confirm construct integrity*
pgl3b-Brn3c-3.6 (p-Luc-1391C)	15	2	SNPG1 [(G)$_8$CG]	C	20	C	Human PAC library RPCI1 (HGMP, MRC, U.K.)		RVprimer3 GLprimer2 3cP1-S
p-Luc-SNPG1$_{-566GT20}$ (p-Luc-1391A)	15	2	SNPG1 [(G)$_8$CG]	A	20	C	p-Luc-1391C by site-directed mutagenesis.	[1]Primers: 3cP1-S and 3cP3-AS followed by *Taq I* digest.	RVprimer3 GLprimer2 3cP2-AS 3cP1-S
p-Luc-SNPG1$_{-566GT16}$	15	2	SNPG1 [(G)$_8$CG]	A	16	C	p-Luc-SNPG1$_{-566GT20}$ by *SacII* digest.	[2]Primers: 3cP5-S and 3cP6-AS.	GLprimer2 3cP4-S
p-Luc-SNPG1$_{-566GT23}$	15	2	SNPG1 [(G)$_8$CG]	A	23	C	p-Luc-SNPG1$_{-566GT20}$ by *SacII* digest.	[2]Primers: 3cP5-S and 3cP6-AS.	GLprimer2 3cP4-S
p-Luc-SNPG2$_{-566GT16}$	21	2	SNPG2 [GGC(G)$_9$]	A	16	C	p-Luc-SNPG1$_{-566GT16}$ by *SacI* and *SpeI* digest.	[3]Primers: 3cP12-S and 3cP12-AS	RVprimer3
p-Luc-SNPG2$_{-566GT23}$	21	2	SNPG2 [GGC(G)$_9$]	A	23	C	p-Luc-SNPG1$_{-566GT23}$ by *SacI* and *SpeI* digest.	[3]Primers: 3cP12-S and 3cP12-AS	RVprimer3

[1] PCR was performed under standard conditions as described in section 3.2.2.1 with 0.2µM of each primer (3cP1-S and 3cP3-AS), 2mM $MgCl_2$ at an annealing temperature of 60.3°C with an annealing and extension stage of 45 seconds. Post-PCR 12µl of the PCR mix was used in a *Taq I* digest under standard conditions (see section 3.2.9.2) to identify the wild-type allele, A (the wild type allele produces 3 bands of 81, 123 and 392bp; the variant allele 2 bands of 204 and 392bp).

[2] PCR was performed under standard conditions (see section 3.2.2.1) with 0.2µM of each primer (3cP5-S and 3cP6-AS), 2mM $MgCl_2$, 5% DMSO at an annealing temperature of 54°C. Only constructs carrying the *SacII* insert in the correct (forward) orientation produce a 490bp product upon PCR.

[3] PCR was performed under standard conditions (see section 3.2.2.1) with 0.2µM of each primer (3cP12-S and 3cP12-AS), 2mM $MgCl_2$ at an annealing temperature of 59°C. Only constructs carrying the *SacI* / *SpeI* insert produce a 400bp product upon PCR. * See Table 3.3 for information on sequencing primers.

Table 3.5: Summary of pGL3-promoter constructs to detail restriction enzymes used to confirm copy number of inserts.

Oligonucleotide name*	Tm (°C)	Hybridisation temperature (°C)	Double restriction enzyme digest screen	pGL3- promoter construct formed**
SNPG1	75.57	71	*Nhe I, Nco I*	pGL3-P(SNPG1)$_2$
SNPG2	77.46	71	*Nhe I, Nco I*	pGL3-P(SNPG2)$_2$
(G)$_{10}$	75.57	71	*Nhe I, Nco I*	pGL3-P(G10)$_2$
(G)$_{11}$	76.54	71	*Nhe I, Nco I*	pGL3-P(G11)$_2$
SP1	74.05	70	*Nhe I, Nco I*	pGL3-P(SP1)$_2$
SP1 Mut	70.90	66	*Nhe I, Nco I*	pGL3-P(SP1Mut)$_2$
1391A	75.00	71	*Nco I, Sac I*	pGL3-P(1391A)$_2$
1391C	78.60	71	*Nco I, Sac I*	pGL3-P(1391C)$_2$
Gfi-1	70.22	66	*Nco I, Sac I*	pGL3-P(Gfi-1)$_2$
Gfi-1 Mut	71.33	66	*Nco I, Sac I*	pGL3-P(Gfi-1Mut)$_3$

* Oligonucleotide sequence and 5'-overhang is given in Table 3.2.

** Lower case denotes copy number of oligonucleotide insert.

4.0 Identification And Characterisation Of Genetic Variation At The Brn-3c Locus.

4.1 Introduction.

The overall aim of this project is to investigate whether common sequence variants in the Brn-3c gene, a hair cell POU domain transcription factor, underlie susceptibility to late onset sensorineural hearing loss (a discussion of the term late onset hearing loss is reviewed in section 1.5.1). Late onset hearing loss is an extremely common and complex trait. Audiometrically, it is characterised by an elevation of hearing thresholds in the high frequencies and at the cellular level a loss of hair cells in the cochlea is widely regarded to underlie the pathology (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller et al, 2004; Gates and Mills, 2005).

At the onset of this project Brn-3c was one of the very few genes that had emerged as a good candidate gene for susceptibility to late onset hearing loss. During the course of this project several additional promising candidate genes for susceptibility to late onset hearing loss have come to light including Cdh23 (involved in stereocilia integrity), Barhl1 (involved in hair cell survival), PMCA2 (involved in maintenance of calcium homeostasis in the cochlea) and some genes involved in regulation of ROS homeostasis in the cochlea (see section 1.7). However, Brn-3c still remains one of the strongest candidates. Brn-3c is one of three highly homologous genes that constitute a small gene family, the Brn-3 family that is categorised under division IV of the widely characterised POU protein family (for review see Latchman, 1999). The additional members of the Brn-3 family are important developmental regulators involved in the maintenance and survival of specific neuronal cell sub-types (Erkman et al, 1996; Gan et al, 1996; McEvilly et al, 1996; Xiang et al, 1996; Eng et al, 2001; see section 1.9.2). Within the inner ear Brn-3c is the only Brn-3 factor expressed in hair cells (Erkman et al, 1996; Xiang et al, 1997; Xiang et al, 1998). Importantly, evidence from Brn-3c (-/-) mice suggests that Brn-3c functions in *hair cell maturation* and *maintenance during development*; in the absence of Brn-3c the molecular cues that direct formation of the highly organised stereociliary hair bundle are lost and a lack of Brn-3c results in the progressive degeneration of incompletely differentiated hair cells by apoptosis such that

the total loss of these cells is evident in the sensory epithelia of the inner ear shortly after birth (Erkman et al, 1996; Xiang et al, 1997; Xiang et al, 1998). Moreover, an 8bp deletion in the POU homeodomain of Brn-3c in humans results in adult onset hearing loss, DFNA15 thus suggesting that Brn-3c is essential for *maintenance of hair cell function in the long term* (Vahava et al, 1998). Collectively, these features make Brn-3c a very strong candidate gene for susceptibility to late onset hearing loss (see section 1.8.1 for a comprehensive discussion of the evidence that Brn-3c is a good candidate gene for late onset hearing loss).

Predisposition to late onset hearing loss as with many traits of a common, complex nature is thought to be due to multiple genetic and environmental risk factors and the relative risk contributed by any one factor is expected to be moderate (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller et al, 2004; Gates and Mills, 2005). Specifically, genetic susceptibility to complex traits such as late onset hearing loss is thought to be due to moderate variations in gene expression or function at multiple loci and the DNA sequence variants responsible for such effects are thought to be common in the general population. This concept that common sequence variants (with a rare allele frequency of or greater than 1%) underlie susceptibility to common diseases is the basis of the CD:CV hypothesis (for reviews see Smith and Lusis, 2002; Wright et al, 2003). Therefore, given the paramount role of Brn-3c in hair cell survival it is hypothesized that common inter-individual sequence variants in the Brn-3c gene or within the regulatory regions of this gene that have a subtle affect on function or expression of Brn-3c respectively, could be a risk factor for susceptibility to this disease.

4.1.1 Genetic variation at the Brn-3c locus reported at the onset of this project.

It is hypothesised that common inter-individual sequence variants in the Brn-3c gene predispose to late onset hearing loss. At the onset of this project an 8bp deletion within the coding region of Brn-3c located in the POU homeodomain had been identified in one family, a large Israeli-Jewish family and found to cause adult onset hearing loss DFNA15 with an autosomal dominant mode of progressive late onset hearing loss (Vahava et al, 1998; see section 1.8.1 and 1.8.2). This 8bp deletion in the Brn-3c gene is sufficient to cause adult onset hearing loss manifesting between ages 18-30 years and appears to be rare but it raises the possibility that there may be common sequence

variants within the Brn-3c gene that have a more subtle effect on phenotype and hence, are only phenotypically significant in the later decades of life. Common sequence variants in the Brn-3c gene that predispose to late onset hearing loss, albeit with a subtle effect on phenotype could manifest within the coding region, intron or regulatory regions of the gene. Vahava et al have previously examined the coding region of Brn-3c for sequence variation in their search for a mutation in Brn-3c that is responsible for DFNA15. However, this analysis was restricted to the Israeli-Jewish family and hence, a small number of samples. It gives no information as to the existence of common sequence variants in the Brn-3c coding region that could predispose large numbers of the general population to late onset hearing loss.

At the time of commencing this project minimal information was reported in the public SNP databases for the Brn-3c gene (see Table 4.1) and the international HapMap project which aims to characterise the common patterns of DNA sequence variation in the human genome had not been formed (International HapMap Consortium, 2003; see section 1.6.2.1). At the onset of this project three single nucleotide substitutions were reported within the vicinity of the Brn-3c coding region. An A>C substitution at position -1391 in the Brn-3c promoter (NCBI SNP Cluster ID: rs1368402) and a G>A and C>T substitution at 230 and 424 nucleotides respectively, beyond the translation termination codon in the Brn-3c gene [hereafter referred to as *230G>A (NCBI SNP Cluster ID: rs891969) and *424C>T (NCBI SNP Cluster ID: rs891970) according the guidelines issued by the Human Genome Variation Society www.hgvs.org/mutnomen/; the 3' end of the mRNA transcript for the Brn-3c gene has not been defined in humans]. However, none of these sequence variants had been validated and the significance of these sequence variants with respect to Brn-3c regulation and / or expression was not known. Furthermore, little was known about how Brn-3c is regulated or maintained in a tissue specific manner in sensory hair cells of the inner ear. No functional cis-acting elements had been characterised within the Brn-3c promoter. Had any such regulatory elements been identified, they would have formed an ideal basis within which to search for common sequence variants in the Brn-3c gene that lead to inter-individual variations in regulation and / or expression of Brn-3c.

Therefore, in the absence of known functional cis-acting elements within the Brn-3c promoter or the first intron of the Brn-3c gene, coupled with a lack of comprehensive information regarding genetic variation at the Brn-3c locus at the time of undertaking this project it was decided to perform mutation scanning of the Brn-3c locus in an attempt to identify common sequence variants in this gene. For the purpose

of this project all regions of the Brn-3c gene including the coding region, intron and promoter were considered.

Table 4.1: Genetic variation at the Brn-3c locus reported in NCBI SNP database at the onset of this project

Variant	Position	Nature	NCBI SNP Cluster ID
-1391A>C	Promoter	A >C	rs1368402
***230G>A**	3'-flanking region	G >A	rs891969
***424C>T**	3'-flanking region	C >T	rs891970

4.1.2 Mutation scanning: single strand conformational polymorphism (SSCP) analysis and other methods.

Mutation scanning is a form of mutation detection whereby specific genes are scanned for the presence of novel sequence variants. There are many technologies available to perform mutation scanning to search for novel sequence variants including denaturing gradient-gel electrophoresis (DGGE), chemical cleavage of mismatch (CCM), heteroduplex analysis (HA), denaturing high performance liquid chromatography (dHPLC), single strand conformational polymorphism (SSCP) analysis and direct sequencing (for reviews see Ellis et al, 1998; Nataraj et al, 1999; Xiao and Oefner, 2001; Knapp, 2004; Suh and Vijg, 2005). Each method has its own potential advantages and disadvantages and these must be taken into consideration together with practical considerations such as the size of the gene to scan, the number of samples to be scanned and the availability or cost of equipment when deciding which procedure to use.

The purpose of this project is to identify common sequence variants (with a rare allele frequency of or greater than 1%) in the Brn-3c gene that may predispose large numbers of the ageing population to late onset hearing loss. Rare sequence variants (with a rare allele frequency of less than 1%) although may be functionally significant with respect to Brn-3c in a limited number of individuals are not the focus of this project. Therefore, since only a modest number of DNA samples needed to be scanned at the onset of this project in order to identify common sequence variants in the Brn-3c gene, this was an important factor when deciding upon the mutation scanning methodology.

SSCP analysis originally developed in 1989 by Orita et al, is a very simple and rapid technique that has been widely used for mutation detection in genes responsible for genetic disease (Castellvi-Bel et al, 1999; Majores et al, 2002; Shinahara et al, 2004). A variety of mutations can be detected by SSCP including single nucleotide substitutions, insertions and deletions the former of which are the most common type of DNA sequence variation in the human genome; single nucleotide substitutions with a rare allele frequency of at least 1% are estimated to occur every 200-300bp (for reviews see Salisbury et al, 2003; Wright et al, 2003). At the onset of this project SSCP analysis was considered as one of the most practical, rapid and reliable methods for mutation scanning of the Brn-3c gene. The principle of SSCP analysis will be discussed and compared alongside additional methods for mutation scanning.

SSCP analysis involves the amplification of sample DNA by PCR, denaturation followed by electrophoresis on a non-denaturing polyacrylamide gel. It is based upon the principle that under non-denaturing conditions, the secondary structure of single stranded DNA molecules is directly related to their primary structure, their nucleotide composition. This effectively means that single stranded DNA molecules that differ in sequence by even a single nucleotide have the ability to fold into two different conformations. These "conformers" can manifest in electrophoretic mobility shifts or in the presence or absence of specific bands upon non-denaturing polyacrylamide gel electrophoresis and can be visualised by autoradiography (for reviews see Nataraj et al, 1999; Suh and Vijg, 2005). Classically, the PCR product of an individual heterozygous for a specific sequence variant such as a single nucleotide substitution should generate at least four bands, two with mobility identical to that of the wild-type allele and two that are characteristic of the sequence change (this is illustrated in Fig 4.1). However, it is important to realise that this is not always the case. One sequence of DNA can give rise to more than two conformations termed, iso-conformers. Iso-conformers can arise for many reasons; the two complementary strands of a DNA molecule can each fold into more than one conformer. In addition, the ability of *Taq* DNA polymerase to randomly append the 3' end of double stranded DNA fragments with an extra residue, usually adenine means that single strands that carry the appendage have the ability to fold into different conformers compared to those that lack the appendage (Sugano et al, 1996). In a similar manner the ability of excess primers to interact with amplicons can also induce the formation of iso-conformers (Cai and Touitou, 1993). However, despite this anomaly, in general unincorporated primers are not removed prior to SSCP analysis as presence of excess primers can lead to sharper SSCP bands (Hennessy et al, 1998). It is

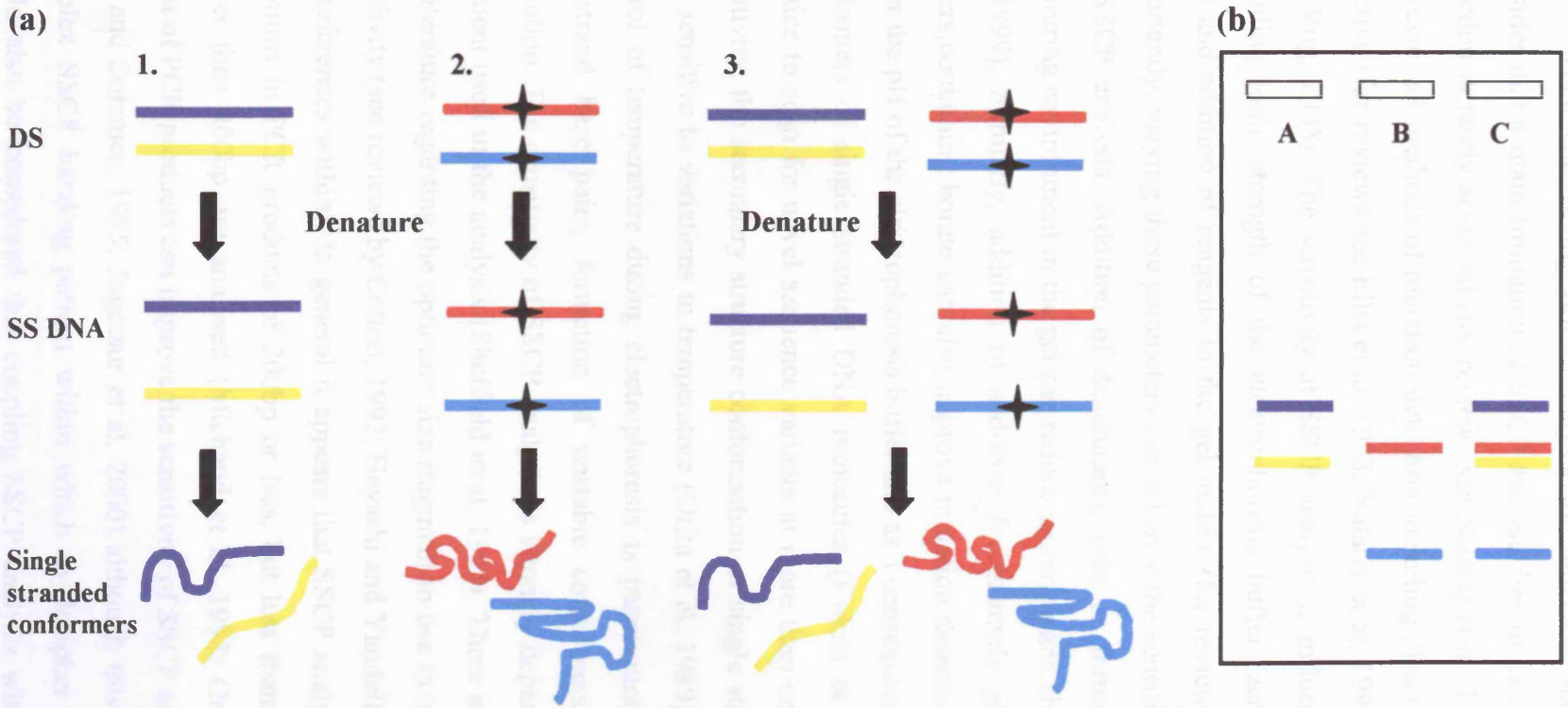


Figure 4.1 A schematic to illustrate the principle of SSCP analysis. (a) Subject 1 homozygous wild type (one allele shown), subject 2 homozygous variant (one allele shown) and subject C heterozygous for the variant allele (both alleles shown). **(b)** SSCP gel for samples 1-3 in (a) to illustrate a classical example of the migration pattern that may be produced by a subject heterozygous for a single sequence variant compared to subjects homozygous wild-type and variant. Four bands are seen for the heterozygous subject, 3, two with mobility identical to that of the wild-type allele and two characteristic of the sequence change.

thought that unincorporated primers enhance the formation of stable conformers most probably by decreasing or abolishing minor or metastable conformations.

When considering SSCP analysis for mutation detection it is important to consider that a main limitation of SSCP analysis lies in its sensitivity; 100% mutation detection is rarely achieved (for reviews see Nataraj et al, 1999; Suh and Vijg, 2005). However, all methods of mutation detection including direct sequencing are not 100% effective (for reviews see Ellis et al, 1998; Nataraj et al, 1999; Franca et al, 2002; Suh and Vijg, 2005). The sensitivity of SSCP analysis is influenced by many parameters including ionic strength of the electrophoresis buffer, acrylamide-to-bis-acrylamide ratio and addition of reagents to the gel matrix (for review see Nataraj et al, 1999). Importantly, varying these parameters can enhance the sensitivity of mutation detection by SSCP analysis. Addition of denaturants, urea or formamide to create a mildly denaturing environment in the gel can reduce smearing and sharpen SSCP bands (Yip et al, 1999). Similarly, addition of additives for example glycerol to electrophoresis buffers containing borate can also improve mutation detection. Glycerol is reported to lower the pH of the electrophoresis buffer and as a consequence stabilises the secondary conformers of single stranded DNA molecules (Kukita et al, 1997). It is common practice to scan for novel sequence variants at more than one temperature to enhance sensitivity; the secondary structure conformation of single stranded DNA molecules is very sensitive to variations in temperature (Orita et al, 1989). For this reason accurate control of temperature during electrophoresis is paramount to avoid dissociation of intrastrand base pairs, formation of unstable conformers and ultimately loss of resolution. The sensitivity of SSCP analysis is strongly dependent upon the size of the fragment used in the analysis (Sheffield et al, 1993). There are conflicting estimates in the literature regarding the optimum size fragment to use in order to achieve maximum sensitivity (see reviews by Cotton, 1992; Hayashi and Yandell, 1993; Nataraj et al, 1999 and references within). In general it appears that SSCP analysis can detect 70-95% of mutations in PCR products of 200bp or less, but less than 50% when fragments of greater than 400bp are analysed (Michaud et al, 1992; Grompe, 1993). Restriction digest of PCR products can improve the sensitivity of SSCP analysis for large fragments (Liu and Sommer, 1995; Jugessur et al, 2000) although this can sometimes lead to a complex SSCP banding pattern within which to decipher abnormal conformers. It should also be considered that coupling SSCP analysis with HA analysis, a related technique, has long been known to increase the sensitivity of mutation detection by SSCP analysis (Ravnik-Glavac et al, 1994; Rossetti et al, 1995).

HA analysis is based upon the conformation of duplex DNA fragments in native gels. It exploits the fact that samples heterozygous for a given allele can form heteroduplexes upon denaturing and re-annealing; for a heterozygous sample four duplex DNA species should form: two perfectly matching homoduplexes (one wild-type the other variant) and two heteroduplexes containing a mismatch (formed by annealing of either the wild-type sense strand with the variant anti-sense strand or vice-versa). To facilitate formation of heteroduplexes in samples homozygous for the rare allele of any given sequence variant samples to be analysed are mixed with wild-type DNA as a reference. Heteroduplex DNA fragments containing a mis-match generally exhibit a retarded electrophoretic mobility compared to the corresponding homoduplex DNA fragments formed by the annealing of complementary strands (for review see Nataraj et al, 1999). During SSCP analysis single stranded DNA fragments denatured post-PCR prior to loading on SSCP gels often renature to a certain extent. Hence, if this occurs in a heterozygote sample heteroduplexes can form between wild-type and mutant strands; these heteroduplexes can often be resolved from the homoduplexes in a lower position from the single stranded conformers on SSCP gels and facilitate the detection of sequence variants (Gerrard and Dean, 1997). When used as the sole means of mutation scanning the sensitivity of HA analysis is similar to that of SSCP analysis (for reviews see Grompe 1993; Nataraj et al, 1999) but a drawback of HA analysis compared to SSCP analysis is that sensitivity appears greater for detection of insertions and deletions compared to base substitutions (Wikman et al, 2000).

Similar to HA analysis dHPLC and DGGE are two additional mutation scanning techniques that exploit inherent properties of heteroduplex DNA fragments compared to homoduplexes; both DGGE and dHPLC exploit the differential denaturation kinetics of hetero- compared to homo-duplex DNA fragments. In DGGE double stranded DNA fragments are electrophorised through a gradient gel of increasing denaturant concentration or temperature (temperature gradient gel electrophoresis, TGGE). This effectively causes the double stranded DNA fragments to denature in a step-wise fashion as discrete melting domains unfold, the domains of low melting temperature denaturing first. A single difference in nucleotide composition between two double stranded DNA fragments is sufficient to alter the denaturation kinetics for the melting domain within which the variant nucleotide resides such that altered electrophoretic mobilities are observed (for reviews see Knapp, 1994; Suh and Vijg, 2005). The sensitivity of DGGC is enhanced by applying heteroduplexing and by attachment of GC-rich clamps to the fragments under analysis the later of which ensures detection of

sequence variants occurs within regions of relatively low melting temperature. Compared to SSCP analysis DGGC is a highly sensitive method, but it is arguably technically more difficult to perform and thorough optimisation of the conditions for gel electrophoresis is required (Suh and Vijg, 2005).

dHPLC utilises ion-pair reverse phase liquid chromatography to differentially retain homo- and hetero-duplex DNA fragments; the heteroduplex DNA fragments elute more rapidly from the column due to their ability to denature more readily (for reviews see Xiao and Oefiner, 2001; Suh and Vijg, 2005). One of the main advantages of dHPLC is that complete automation can be achieved using the WAVE DNA fragment analysis system, which is adapted for 96-well throughput. Samples are detected by on-line UV absorbance at 260nm and can be visualised as an elution profile; heterozygous samples typically displaying four chromatographic peaks: two representing the homoduplexes and two representing the heteroduplexes (Donohoe, 2004). However, current drawbacks of this system compared to SSCP analysis are that only one sample can be analysed at a time resulting in a modest throughput. In addition, the operating temperature has to be optimised for each fragment to be assayed and initial set-up costs are high due to the need to purchase specialist equipment (for review see Suh and Vijg, 2005).

SSCP analysis is limited in that it does not define the nature of the sequence variant or the location of the sequence variant within the fragment examined. However, this anomaly is not limited to SSCP analysis many methods of mutation scanning including HA, DGGE and dHPLC pin-point putative sequence variants without identifying the position or nature of the sequence variant within the fragment of DNA under analysis. Therefore, abnormal bands identified by SSCP or these alternative mutation scanning techniques must be followed up by sequencing analysis in order to identify the sequence change. Indeed, direct sequencing is considered the 'Gold standard' amongst mutation detection techniques because it defines precisely the nature and location of the sequence change (for review see Suh and Vijg, 2005). However, direct sequencing, as the sole means of mutation scanning may not be the most cost-effective approach. Even with the use of the current genetic analysers that utilise capillary arrays and fluorescent dye deoxy terminator methodology for example, the ABI Prism 3100 Genetic Analyser fitted with a standard 50cm capillary array and used in combination with BigDye terminator chemistries (Applied Biosystems) mutation scanning of large, complex, multi-exon genes by direct sequencing can be very time

consuming and costly, especially if a large number of samples have to be processed (for reviews see: Franca et al, 2002; Suh and Vijg, 2005).

Besides direct sequencing, one of the few mutation scanning techniques that localises the sequence variant within the fragment under examination is CCM. CCM was originally described by Cotton in 1988 and similar to HA, dHPLC and DGGE, utilises heteroduplex formation between wild-type and variant DNA fragments. Typically, the wild-type DNA is given a label (a radiolabel or fluorescence label) and added to all sample DNA to be scanned. Heteroduplex formation is followed by chemical modification at the sites of the mismatched bases; osmium tetroxide is used to chemically modify mismatched thymines and hydroxylamine for mismatched cytosines. Piperidine is then used to cleave the modified strand at the site of the mismatch and cleavage products detected by denaturing gel electrophoresis followed by autoradiography or fluorescence detection, as appropriate (for review see Ellis et al, 1998). However, drawbacks of the CCM method compared to SSCP are the use of toxic chemicals and it is technically more difficult to perform. CCM is probably most useful when large and complex genes have to be examined as fragments over 1Kb can be screened at a time (for review see Ellis et al, 1998).

SSCP analysis is a popular choice for mutation scanning and the main advantages of SSCP analysis particularly when compared to dHPLC, DGGE and CCM most probably lie in its speed and / or simplicity. The use of pre-cast gels that can examine up to 48 samples at a time (Pharmacia Biotech) and the application of silver staining to visualise SSCP bands eliminates the need for radioactivity and has been reported to increase resolution (Dockhorn-Dwornickzak et al, 1991). Enhancements in SSCP analysis over recent years have lead to the incorporation of fluorescent labels and the separation of fragments by automated capillary array electrophoresis using automated genetic analysers such as the ABI PRISM 3100 (PE Applied Biosystems); a procedure which has greatly streamlined the mutation detection process and increased the turn-around time (Baba et al, 2003; for review see Suh and Vijg, 2005).

Finally, attention should be drawn to the fact that SSCP analysis as with many techniques used in mutation detection (eg: dHPLC, DGGE, HA and CMC) is based on PCR amplification of sample DNA prior to analysis. Consequently, it is very important to validate reports of novel mutations that have been obtained via these techniques particularly the identification of single nucleotide substitutions. *Taq* DNA polymerase introduces errors into the newly synthesized strand during PCR and does not have 3'-5' exonuclease proofreading activity; Cline et al, 1996 report an error rate of 8.0×10^{-6} for

Taq DNA polymerase (expressed as mutation frequency/bp/duplication). Therefore, if a sequence change is only detected in one individual, this should always be confirmed in a second PCR from the same individual to rule out the possibility of a PCR artefact.

In summary, given the small size of the Brn-3c gene (in humans the coding region spans 1017bp; NCBI GenBank Accession: NM_002700) coupled with the fact that only a modest number of samples needed to be scanned it was decided that SSCP analysis in combination with non-radioactive silver staining would be an ideal mutation scanning method with which to examine the Brn-3c gene for the presence of common sequence variants. The relatively high sensitivity of SSCP analysis when more than one condition of electrophoresis is used and the relative ease and speed of the method plus the low-reagent cost made mutation scanning of the Brn-3c gene by SSCP a rapid, reliable, cost-effective and most appealing approach.

4.1.3 Reporting DNA sequence variation: nomenclature.

DNA sequence variation can manifest in various forms including single nucleotide substitutions, deletions or insertions of one, several or even hundreds of base pairs and simple sequence repeats (SSRs) which are tandem repeats of nucleotide motifs commonly referred to as micro- or mini- satellites depending on the length of the repeating motif. Single nucleotide substitutions are the most common type of DNA sequence variation in the human genome, they are mainly located within non-coding regions and it is estimated that there are greater than 10 million single nucleotide substitutions with rare allele frequencies above 1% (for review see Wright et al, 2003). Single nucleotide substitutions are often referred to in the literature as single nucleotide polymorphisms (SNPs). However, this later term is misleading. The classical definition of a polymorphism states a polymorphism is: "a Mendelian trait that exists in the population in at least two phenotypes, neither of which occurs at a frequency less than 1%" (Vogel and Motulsky, 1986) and hence, it has an effect on phenotype. However, the term polymorphism is also used to describe a neutral change in the DNA sequence that has no affect on phenotype and is often used to simply refer to a change in the DNA sequence present at a frequency of 1% or higher in the population regardless of whether it may have an affect on phenotype (see Cotton and Scriver, 1998). Clearly, many of these terms are not well defined and are used ambiguously in the literature. A polymorphism can have a functional affect on gene regulation and / or function of the encoded protein and hence, be of phenotypic significance or it can be neutral without

phenotypic effect. Therefore, for the purpose of this thesis the term *'sequence variant'* as opposed to 'polymorphism' was adopted as the nomenclature of choice when reporting sequence variation identified at the Brn-3c locus and knowledge of the frequency and functional effect(s) of the sequence variant was unknown. This concept is in agreement with the current guidelines set down by the Human Genome Variation Society (www.hgvs.org/mutnomen/). Where the term polymorphic or polymorphism is used within the text in this thesis this represents a variable DNA element and no indication of functional significance is intended to be inferred; it could be neutral or disease causing.

4.2 The PCR-SSCP screen of the Brn-3c gene.

To search for common sequence variants at the Brn-3c locus mutation scanning by PCR-SSCP analysis was performed using DNA isolated from 45 random individuals (see method section: 3.2.1, 3.2.2.1 and 3.2.3). In humans the coding region of the Brn-3c gene is contained within two exons of 120bp (NCBI GenBank Accession: U10060.1) and 897bp (NCBI GenBank Accession: U10061.1) respectively, intervened by a small intron of 315bp (NCBI GenBank Accession: AF043452.1). The small size of this gene facilitated rapid analysis by PCR-SSCP analysis. To enhance the sensitivity of SSCP analysis PCR primers were designed to generate fragments that on average spanned 250bp (see Table 3.1, section 3.2.18). In the quest to achieve maximum sensitivity PCR primers were placed to generate fragments in any given region that overlapped as much as was possible, so that segments of DNA were examined in the context of more than one PCR product as much as was feasible within the limitations of the DNA sequence on primer design. It is particularly important to consider searching for novel sequence variants in the context of more than one PCR product if possible and / or practical as if a sequence variant exists in one allele at the site of primer annealing then this allele may not be amplified in the PCR pre SSCP analysis. Accordingly this sequence variant would go undetected in heterozygous samples. Furthermore, sequence variants located near the end of DNA fragments but not within the sites of primer annealing are not as well detected as those within or closer to the middle of the fragment under analysis (Rossetti et al, 1995).

A total of 14 overlapping PCR fragments were examined at the Brn-3c locus: 6 covering the coding region and intron of the gene and 8 spanning 1.5Kb of the 5'-flanking region (Fig. 4.2). An additional fragment designed to include the dinucleotide $(GT)_n$ repeat at position -3495bp within the 5'-flanking region of the gene was also examined (Fig. 4.2). The PCR was optimised by varying the magnesium concentration and the annealing temperature for each of the 15 PCR reactions to ensure specific amplification of the required fragment for subsequent SSCP analysis. Specificity of the PCR analysis was confirmed by electrophoresing an aliquot from each sample to be analysed on an agarose gel. In order to stand the best chance of detecting all sequence variants at the Brn-3c locus within the region under examination the SSCP screen was performed at three different temperatures of 4, 10 and 20 °C for each fragment (see method section 3.2.3). Samples displaying abnormal conformers in the SSCP screen indicative of putative sequence variants were sub-cloned into the pGEM®-T easy vector

(Promega) alongside samples that did not display the abnormal conformers. The pGEM®-T easy vector has single 3'-thymidine overhangs at the multiple cloning site that facilitates insertion of PCR products with adenosine appendages at their 3'-ends. Several clones were sequenced for each sample in both directions on an ABI 3100 genetic analyser and compared to the reference sequence (for reference sequence see Appendix B) to confirm and to identify the nature of the variation. Using this approach five novel sequence variants were identified at the Brn-3c locus and the presence of two previously reported sequence variants was confirmed. Identification of these sequence variants is discussed in the following sections: 4.3 – 4.4.

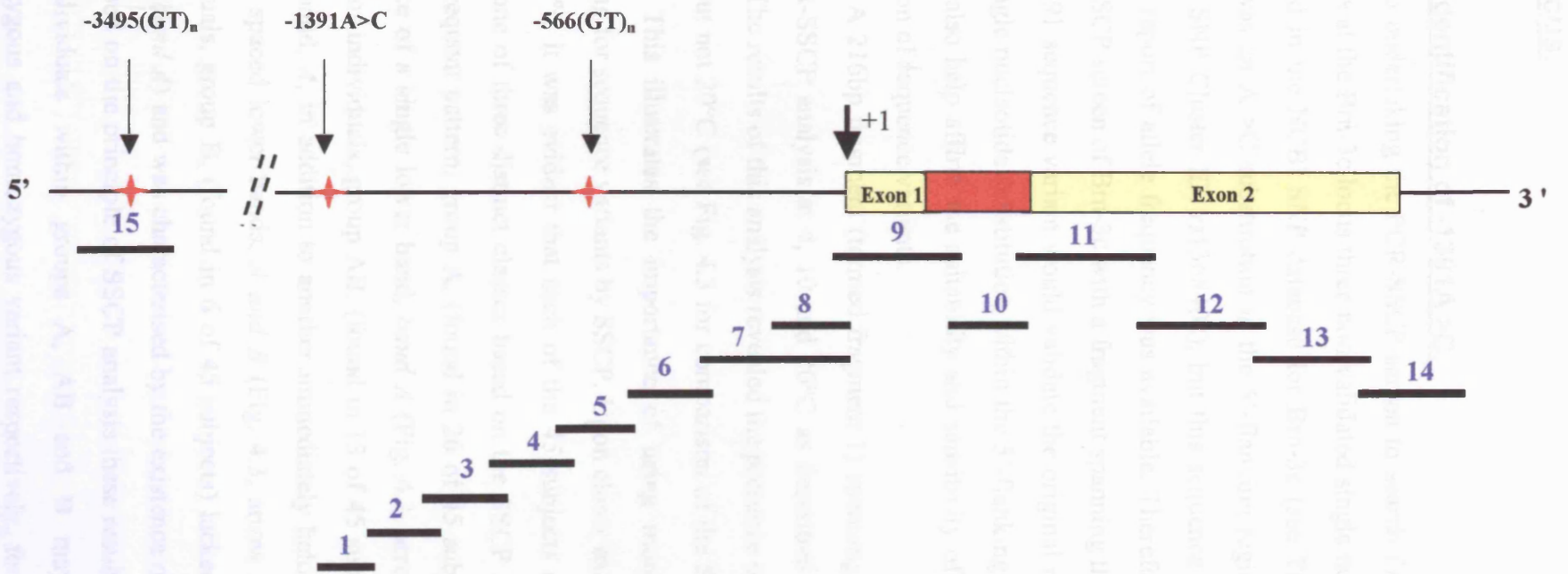


Figure 4.2 Schematic of PCR fragments designed for SSCP analysis at the Brn-3c locus. Exons 1 and 2 are numbered and the intron is shown as a red shaded box (not to scale). The position of the previously reported but non-validated -1391A>C single nucleotide substitution (NCBI SNP Cluster ID: rs1368402) and the location of two dinucleotide (GT) repeats at -566 and -3495 are all indicated by arrows. Numbering is with respect to the first A of the ATG translation start site as +1 (indicated by a small black arrow). Fragments generated by PCR for SSCP analysis are numbered in blue and their locations defined by the thick black lines below. Sequences of primers for each fragment are shown in Table 3.1 Method section 3.2.18. For simplicity the region of the Brn-3c promoter between -1391 and -3495 has been omitted as indicated by the two hatched lines.

4.3 Identification of single nucleotide substitutions at the Brn-3c locus.

4.3.1 Identification of -1391A >C.

Prior to undertaking the PCR-SSCP screen to search for novel and common sequence variants at the Brn-3c locus three non-validated single nucleotide substitutions had been reported in the NCBI SNP database for Brn-3c (see Table 4.1 section 4.1.1). One of these was an A >C substitution in the 5'-flanking region of Brn-3c at position -1391 (NCBI SNP Cluster ID: rs1368402), but this sequence variant had not been validated and no report of allele frequency was available. Therefore, it was decided to begin the PCR-SSCP screen of Brn-3c with a fragment spanning the -1391 locus. Confirmation of the -1391 sequence variant would validate the original report and confirm the presence of a single nucleotide substitution within the 5'-flanking region of Brn-3c. This analysis would also help affirm the suitability and sensitivity of this PCR-SSCP screen for the detection of sequence variants.

A 216bp fragment (termed fragment 1) spanning the -1391 locus was subjected to PCR-SSCP analysis at 4, 10 and 20°C as described in method section 3.2.2.1 and 3.2.3. The results of this analysis revealed the presence of abnormal conformers at 4 and 10°C but not 20°C (see Fig. 4.3 for comparison of the SSCP banding pattern at 10 and 20°C). This illustrates the importance of using more than one temperature when scanning for sequence variants by SSCP. Upon closer examination of the SSCP gels at 4 and 10°C it was evident that each of the 45 subjects examined could be categorised under one of three distinct classes based on the SSCP banding pattern (Fig. 4.3). The most frequent pattern, group A, (found in 26 of 45 subjects) was characterised by the presence of a single lower band, *band A* (Fig. 4.3, arrow A). This was followed by a subset of individuals, group AB, (found in 13 of 45 subjects) that displayed this single lower band, *A*, in addition to another immediately below, *band B*, hence forming two closely spaced lower bands, *A and B* (Fig. 4.3, arrow AB). Finally a small subset of individuals, group B, (found in 6 of 45 subjects) lacked the upper of these two lower bands (*band A*) and was characterised by the existence of *band B* alone (Fig. 4.3, arrow B). Based on the principle of SSCP analysis these results suggest that the genotypes of the individuals within groups A, AB and B may be homozygous wild-type, heterozygous and homozygous variant respectively, for a particular sequence variant. Indeed, for a common sequence variant one would generally expect individuals that are

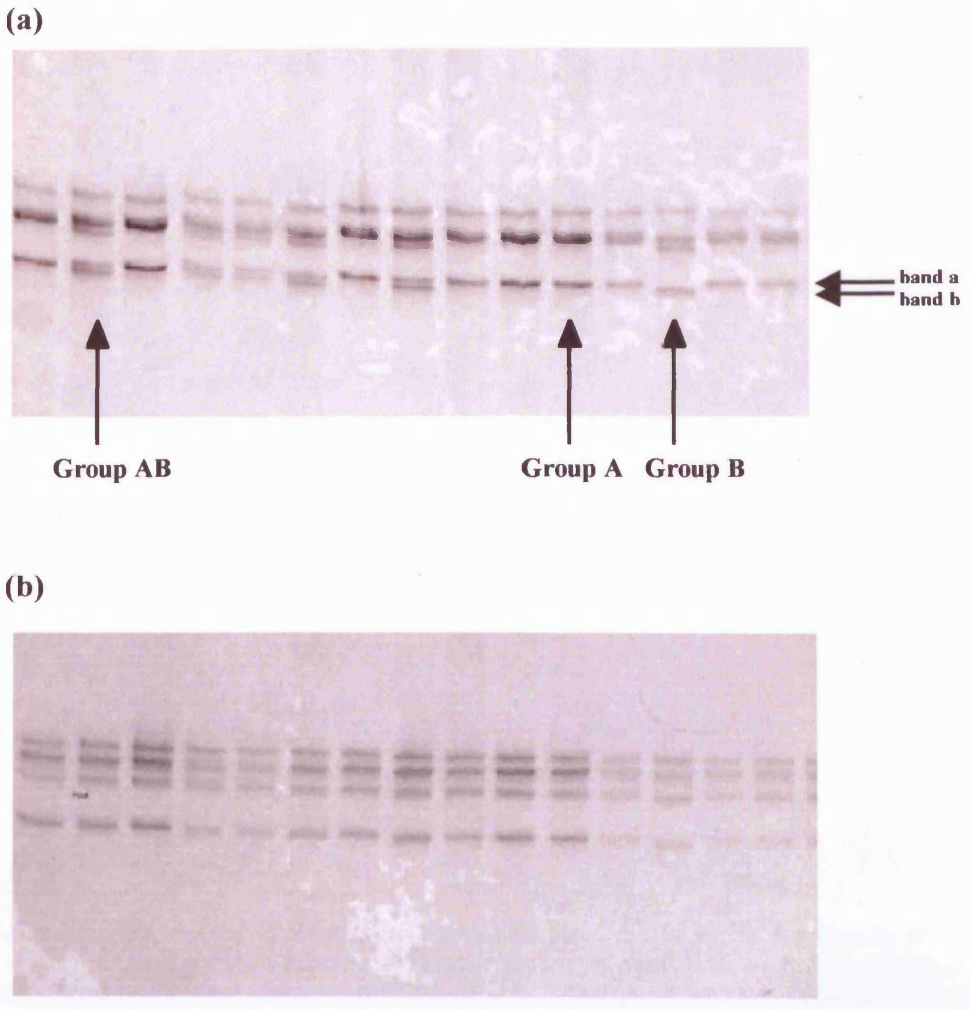


Figure 4.3 SSCP results for fragment 1 spanning the -1391 locus in the Brn-3c 5'-flanking region. (a) Section of the SSCP gel electrophoresed at 10°C compared to the same section in **(b)** electrophoresed at 20°C. Results are shown for 15 individuals. The presence of abnormal conformers can be observed at 10°C but not 20°C. Subjects can be categorised into three distinct groups shown by arrows A, B, and AB based on the SSCP banding pattern observed in (a). Bands a and b are indicated. Primer sequences for fragment 1 are shown in Table 3.1 Method section 3.2.18.

homozygous wild-type to exhibit the most frequent banding pattern and individuals homozygous for the variant allele to exhibit the least frequent banding pattern with individuals that are heterozygous exhibiting a banding pattern at frequency intermediate between the two. However, as discussed previously (section 4.1.2) a limitation of SSCP analysis is that it does not define the nature of the sequence variant or the location of the sequence variant within the fragment under analysis. Therefore, SSCP is not an accurate method by which to assign genotypes. Indeed, two distinct single nucleotide substitutions at the same position can result in identical SSCP shifts; different base substitutions at any particular nucleotide position do not necessarily mean that distinct SSCP migration patterns will be observed (Sheffield et al, 1993). Therefore, to identify the sequence change(s) within fragment 1 and to assess whether the -1391 A > C

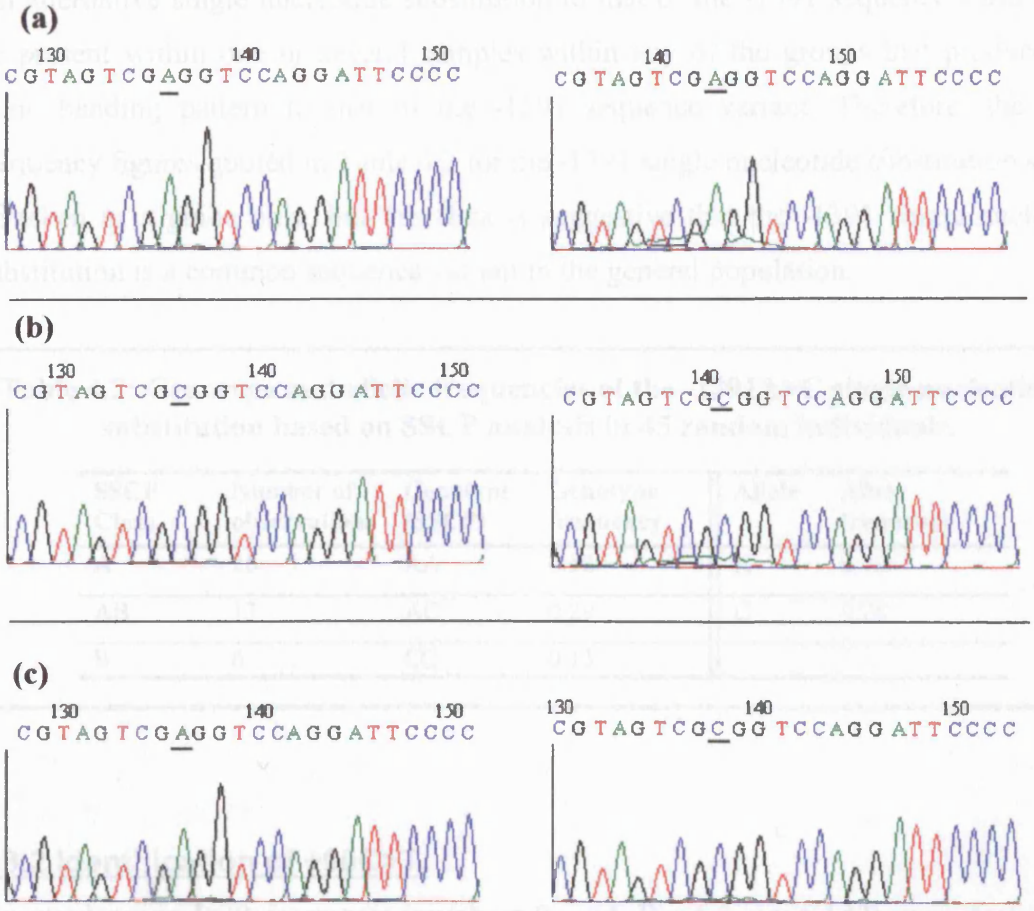


Figure 4.4 Sequencing results for fragment 1 spanning the -1391 locus in the Brn-3c 5'-flanking region. Representative sequencing chromatograms obtained from sequencing DNA cloned from a single individual within each of the three SSCP groups A in **(a)**, B in **(b)** and AB in **(c)**. The -1391 A>C substitution was confirmed as the only sequence change in fragment 1 and the genotypes of the individuals within groups A **(a)**, B **(b)** and AB **(c)** were verified as homozygous wild-type A/A, homozygous variant C/C and heterozygous A/C, respectively. The single nucleotide substitution is underlined in black in each chromatogram. Numbers above each sequencing chromatogram represent the automatic base numbering from the start of readable sequence using BioEdit Sequence Alignment Editor software.

substitution is responsible several clones of an individuals DNA within pGEM®-T easy vector were sequenced for one individual within each of the three groups: A, B and AB. The results of this analysis confirmed the -1391 substitution as the only sequence change within fragment 1 (see Fig. 4.4). In addition, the genotypes of the individuals within groups A, B, and AB were verified as homozygous wild-type A/A, homozygous variant C/C and heterozygous A/C, respectively. This is consistent with the predicted genotypes based on SSCP banding patterns. Based on this data one could estimate that the -1391 sequence variant is a common sequence variant with a rare allele frequency of 0.28 (see Table 4.2). However, as discussed previously SSCP is not a reliable method with which to perform genotyping. Without verifying the genotypes of each sample within a particular group by a direct means for example, sequencing analysis one cannot conclude equivocally that each sample within a particular group has the same genotype.

An alternative single nucleotide substitution to that of the -1391 sequence variant may be present within one or several samples within any of the groups that produces the same banding pattern to that of the -1391 sequence variant. Therefore, the allele frequency figures quoted in Table 4.2 for the -1391 single nucleotide substitution should be taken as a guide only, but this data is suggestive that the -1391 single nucleotide substitution is a common sequence variant in the general population.

Table 4.2: Genotype and allelic frequencies of the -1391A>C single nucleotide substitution based on SSCP analysis in 45 random individuals.

SSCP Class	Number of observations	Genotype (SSCP)	Genotype frequency	Allele	Allele frequency
A	26	AA	0.58	A	0.72
AB	13	AC	0.29	C	0.28
B	6	CC	0.13		

4.3.2 Identification of +90C>T.

Six overlapping PCR fragments (numbers 9 – 14, Fig 4.2, page 142) were designed to examine the complete coding region of Brn-3c and the only intron for the presence of novel sequence variants. Each of these fragments was examined by PCR-SSCP analysis at 4, 10 and 20°C as described in method section 3.2.2.1 and 3.2.3. The results of this analysis showed the presence of abnormal conformers for fragment nine only (Fig. 4.5). Fragment nine spans the whole of exon 1 of the Brn-3c gene (Fig. 4.2, page 142). SSCP analysis of fragment nine at 20°C revealed the presence of abnormal conformers in the form of an extra upper band in two of the 45 subjects examined (Fig. 4.5, arrows A and B). The remaining subjects all conformed to the same banding pattern that lacked the presence of this extra band (Fig. 4.5, comparing arrows A and B with C). Abnormal conformers were not evident for fragment nine when examined at 4 or 10°C.

To identify the sequence change(s) within fragment nine several clones obtained from cloning this fragment into pGEM®-T easy vector from both individuals that displayed the extra conformer (Fig. 4.5, subjects indicated by arrows A and B) were sequenced and compared with clones from one individual who did not display the presence of the abnormal conformer (Fig. 4.5, subject indicated by arrow C). The results of this analysis identified a C to T substitution at position +90 in subjects displaying the abnormal conformer; these subjects were both confirmed by sequencing

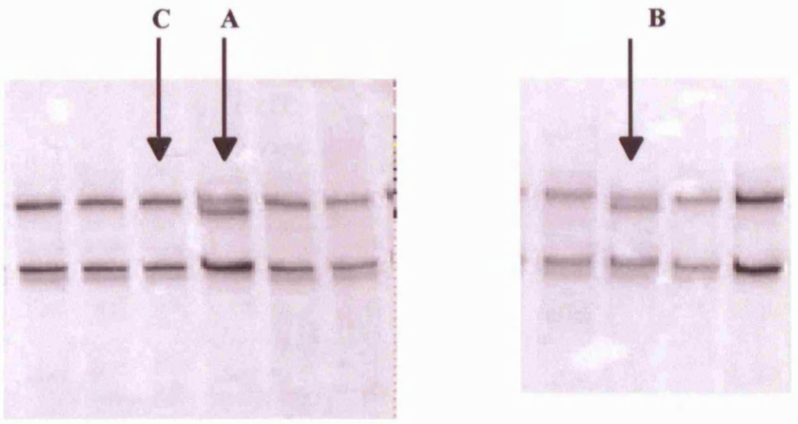


Figure 4.5 SSCP results for fragment 9 spanning exon 1 of Brn-3c. Sections of the SSCP gel electrophoresed at 20°C to show the two subjects containing the abnormal conformer (arrows A and B). All remaining subjects displayed the same banding pattern as indicated by arrow C. Primer sequences for fragment 9 are shown in Table 3.1 Method section 3.2.18.

analysis as heterozygous for the C to T substitution (Fig. 4.6a and b). The subject who did not display the abnormal conformer (indicated by arrow C, Fig. 4.5) was determined homozygous wild type, C/C (Fig. 4.6c). Further analysis revealed that the +90C>T substitution is a synonymous variant; both wild-type (GCC) and variant alleles (GCT) encode for the amino acid alanine (Fig. 4.6). This sequence variant is discussed further in section 4.5.

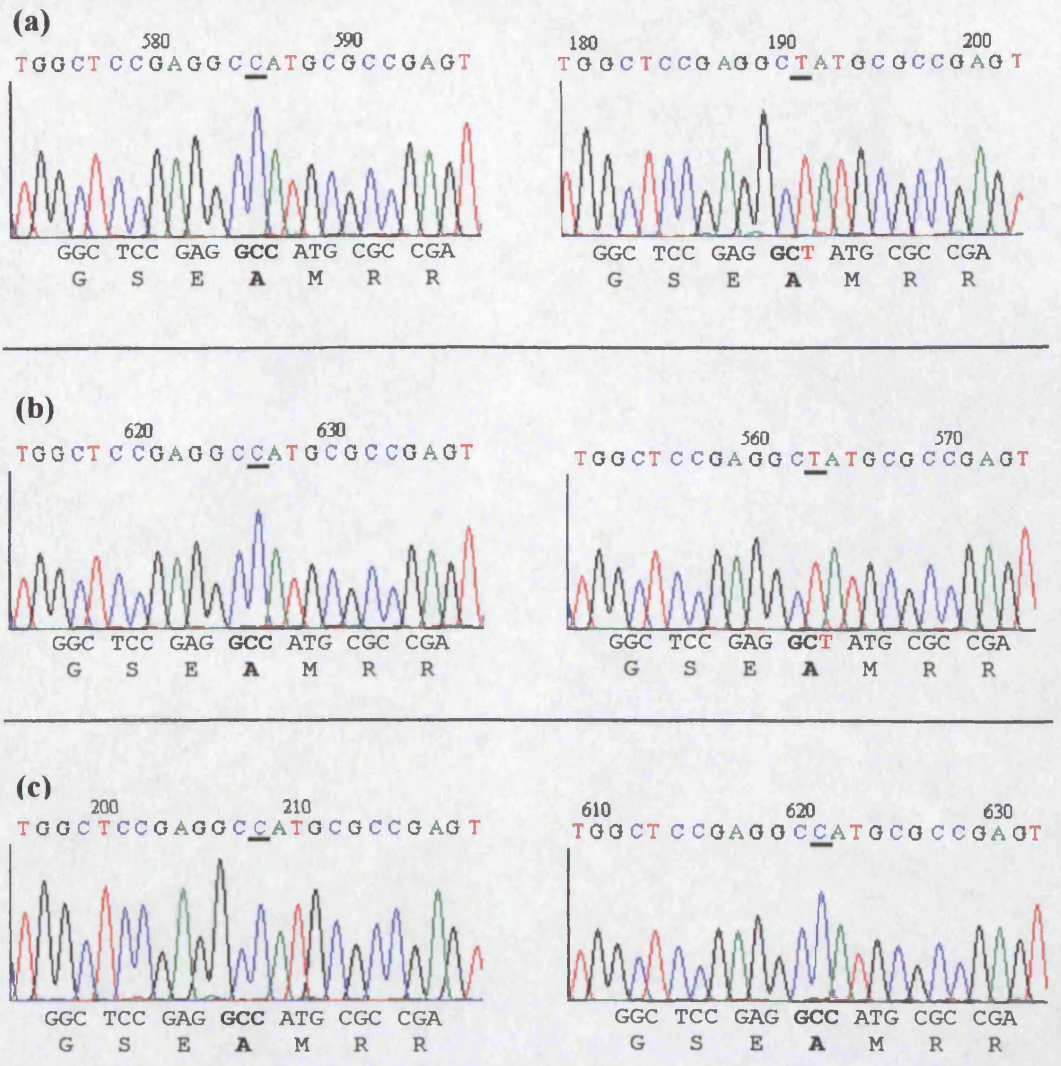


Figure 4.6 Sequencing results for fragment 9 spanning exon 1 of Brn-3c. All sequencing chromatograms were obtained from sequencing cloned PCR products in pGEM®-T easy vector. **(a)** and **(b)** representative sequencing chromatograms for subjects heterozygous for the +90C>T substitution (subjects displaying the extra conformer as indicated by arrows A and B in Fig. 4.5). **(c)** Representative sequencing chromatograms for a subject homozygous wild-type, C/C (subject lacking the extra conformer as indicated by arrow C in Fig. 4.5). The single nucleotide substitution is underlined in black in each chromatogram. The +90C>T substitution is a synonymous variant as illustrated by the amino acid sequence shown under each chromatogram. Numbers above each sequencing chromatogram represent the automatic base numbering from the start of readable sequence using BioEdit Sequence Alignment Editor software.

4.3.3 Identification of -386C>A.

Eight overlapping PCR fragments (numbers 1 – 8, Fig 4.2, page 142) were designed to examine approximately 1.5Kb of the immediate 5'-flanking region of the Brn-3c gene for the presence of novel sequence variants. Each of these fragments was examined by PCR-SSCP analysis at 4, 10 and 20°C as described in methods section 3.2.2.1 and 3.2.3. The results of this analysis confirmed the previously reported -1391A>C sequence variant within fragment one as discussed in section 4.3.1. In addition, an abnormal

SSCP migration pattern was evident following PCR-SSCP analysis of fragment five (the PCR-SSCP results of fragment five are discussed further in section 4.4: Identification of multi- allelic sequence variants) and fragment six (see Fig. 4.7). None of the additional fragments covering 1.5Kb of the 5'-flanking region of Brn-3c showed an abnormal SSCP migration pattern to suggest the presence of a novel sequence variant(s). SSCP analysis of fragment six at 4 and 10, but not 20°C revealed an abnormal migration pattern in one of the 45 samples examined; this sample displayed an abnormal conformer in the form of an extra lower band that was absent in all the other samples examined (Fig. 4.7, arrow A).

To identify the sequence change(s) within fragment six several clones obtained from cloning this fragment into pGEM®-T easy vector for the individual that displayed the extra conformer (Fig. 4.7, subject indicated by arrow A) were sequenced and compared with clones from one individual who did not display the presence of the abnormal conformer (Fig. 4.7, subject indicated by arrow B). The results of this analysis identified a C to A substitution at position -386 and revealed that the subject displaying the abnormal conformer was heterozygous for this sequence variant (Fig. 4.8a). The subject lacking the abnormal conformer (indicated by arrow B, Fig. 4.7) was determined homozygous wild type, C/C (Fig. 4.8b). Since the -386C>A variant was only identified in one individual this raises the possibility that the -386C>A variant may have risen from a PCR error. To rule out the possibility that the -386C>A variant was an artefact of the PCR amplification the PCR amplification was repeated with DNA from the individual displaying this abnormal conformer. SSCP analysis at 4°C on this second

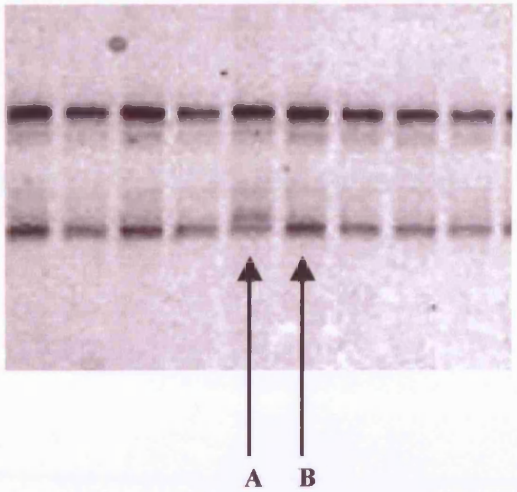


Figure 4.7 SSCP results for fragment 6 within the 5'-flanking region of Brn-3c. Section of the SSCP gel electrophoresed at 4°C to show the subject containing the abnormal conformer (arrow A). All remaining subjects confirmed to the banding pattern as indicated by arrow B. Primer sequences for fragment 6 are shown in Table 3.1 Method section 3.2.18.

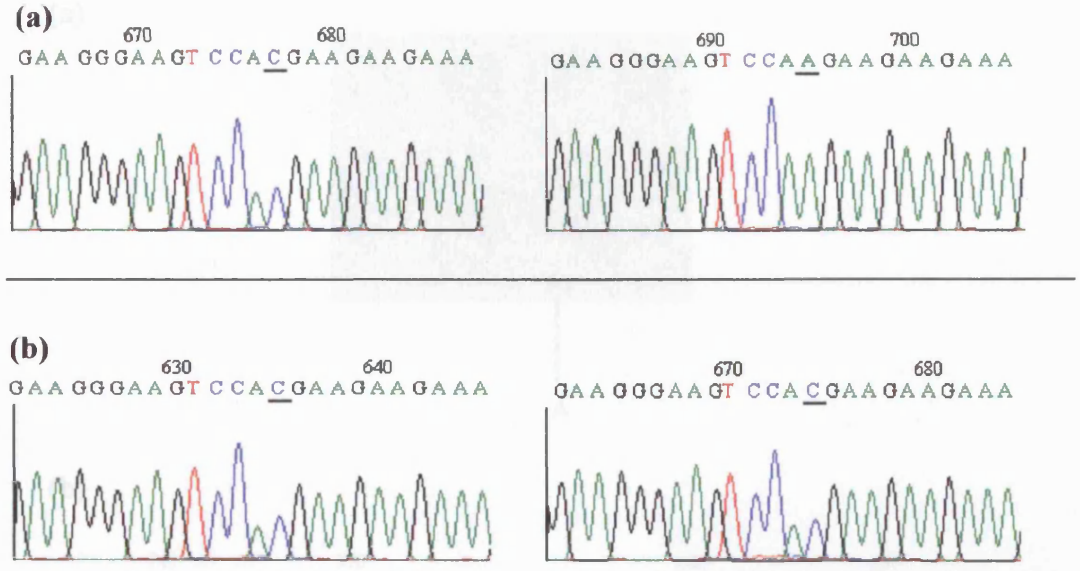
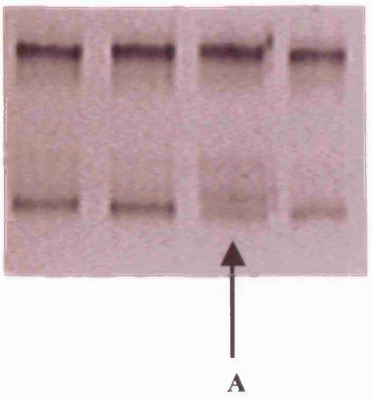


Figure 4.8 Sequencing results for fragment 6 within the Brn-3c 5'-flanking region. All sequencing chromatograms were obtained from sequencing cloned PCR products in pGEM®-T easy vector. **(a)** Representative sequencing chromatograms for the subject heterozygous for the -386C>A substitution (subject displaying the extra conformer as indicated by arrow A in Fig. 4.7). **(b)** Representative sequencing chromatograms for a subject homozygous wild-type, C/C (subject indicated by arrow B in Fig. 4.7). The single nucleotide substitution is underlined in black in each chromatogram. Numbers above each sequencing chromatogram represent the automatic base numbering from the start of readable sequence using BioEdit Sequence Alignment Editor software.

PCR sample of fragment six confirmed the presence of the abnormal conformer in the form of the extra lower band (Fig. 4.9a). Sequencing several clones generated from this second PCR amplification confirmed the C to A substitution at position -386 in the Brn-3c promoter was not an artefact of the PCR amplification (Fig. 4.9b). Collectively, the results of this analysis suggest that the -386C>A substitution is a rare variant; the C to A substitution was only identified in one of the 45 subjects examined and no subjects homozygous for the variant allele were identified from the results of this PCR-SSCP screen. This sequence variant is discussed further in section 4.5.

(a)



(b)

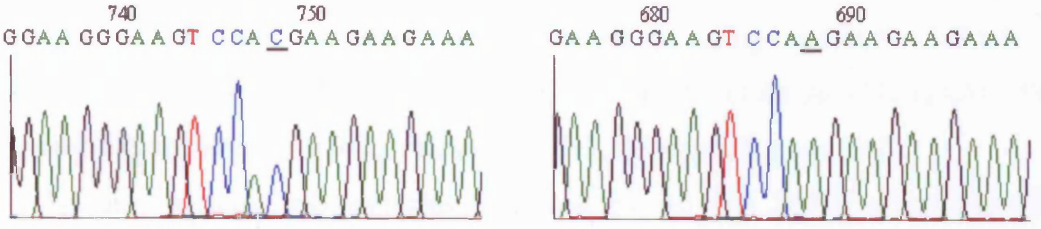


Figure 4.9 Confirmation of -386C >A by repeat PCR-SSCP analysis followed by sequencing. (a) Repeat SSCP results for fragment 6 within the 5'-flanking region of Brn-3c. Section of the SSCP gel electrophoresed at 4°C to show the subject containing the abnormal conformer (arrow A). **(b) Repeat sequencing results for fragment 6 within the Brn-3c 5'-flanking region.** Representative sequencing chromatograms obtained from sequencing cloned PCR products in pGEM®-T easy vector for the subject confirmed by repeat PCR-SSCP analysis to display the extra conformer [subject indicated by arrow A in **(a)**]. Sequencing analysis confirms that the subject is heterozygous for the -386C>A single nucleotide substitution. The single nucleotide substitution is underlined in black in each chromatogram. Numbers above each sequencing chromatogram represent the automatic base numbering from the start of readable sequence using BioEdit Sequence Alignment Editor software.

4.4 Identification of multi-allelic sequence variants at the Brn-3c locus.

4.4.1 Identification of -566$(GT)_n$ repeat variation.

Examination of the 5'-flanking region of Brn-3c using the reference sequence (Ensembl transcript ID: ENST00000230732; Appendix B) revealed a $(GT)_n$ dinucleotide repeat with 21 GT repeats at position -566. Dinucleotide repeats such as these are often polymorphic exhibiting multiple alleles with sequence variation. Moreover, there is increasing evidence that polymorphic dinucleotide repeats have quantitative effects on gene expression (Akai et al, 1999; Gebhardt et al, 1999; Borrmann et al, 2003; Tadokoro et al, 2004). To examine the possibility that the -566$(GT)_n$ repeat is polymorphic PCR-SSCP analysis was performed using primers 3cP5-S: 5' GGACTCTGGTGGACAGCTTG 3' and 3cP5-AS: 5' TGTCCCAGCTCGAACTGCC 3' designed to flank the -566$(GT)_n$ repeat (Fig. 4.2, fragment 5, page 142).

At each of three temperatures examined (4, 10 and 20°C) a variable SSCP banding pattern was observed across the 45 subjects scanned. At 4 and 10°C the SSCP banding pattern was complex consisting of many extra bands. However, three common banding patterns appeared evident at 20°C (Fig. 4.10, as indicated by arrows A, B and C) in addition to a few less frequent banding patterns. These SSCP banding patterns mainly resulted from slight shifts in the migration of specific bands (for example, comparing the banding pattern indicated by arrows B and C in Fig. 4.10). Consequently,

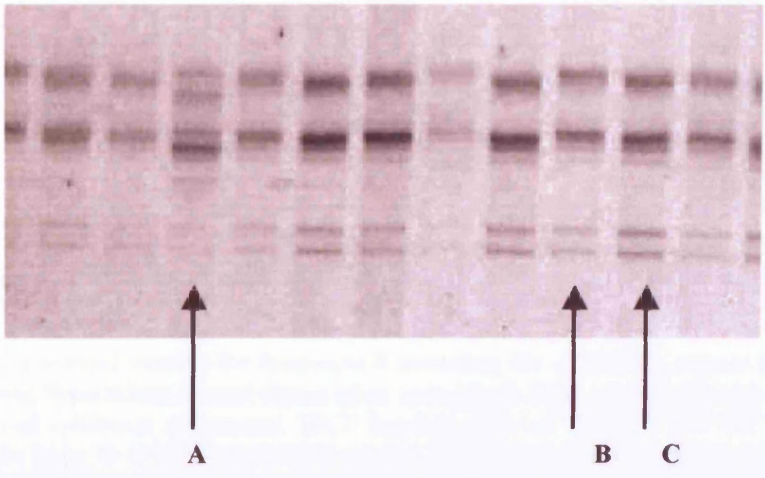


Figure 4.10 SSCP results for fragment 5 spanning the -566(GT)n repeat in the Brn-3c 5'-flanking region. Section of the SSCP gel electrophoresed at 20°C to show the three main banding patterns observed as indicated by arrows A, B and C. Subtle differences in the migration pattern can be observed between samples B and C. Primer sequences for fragment 5 are shown in Table 3.1 Method section 3.2.18.

classifying all the samples into particular subsets as performed for the -1391A>C sequence variant (section 4.3.1) was not considered reliable and so not undertaken. Using the SSCP results obtained at 20°C as a guide, five subjects were selected that clearly exhibited differential SSCP banding patterns. Sequencing several clones of an individuals DNA within pGEM®-T easy vector for each of these five subjects revealed that the -566(GT)n repeat varied in length from 16 to 23 dincucleotide repeats [-566$(GT)_{16\text{-}23}$] (see Fig. 4.11). No additional sequence variants were identified within fragment 5 in 20 clones sequenced. Taken together with the PCR-SSCP results these results suggest that the -566$(GT)_n$ repeat is a common sequence variant in the general population.

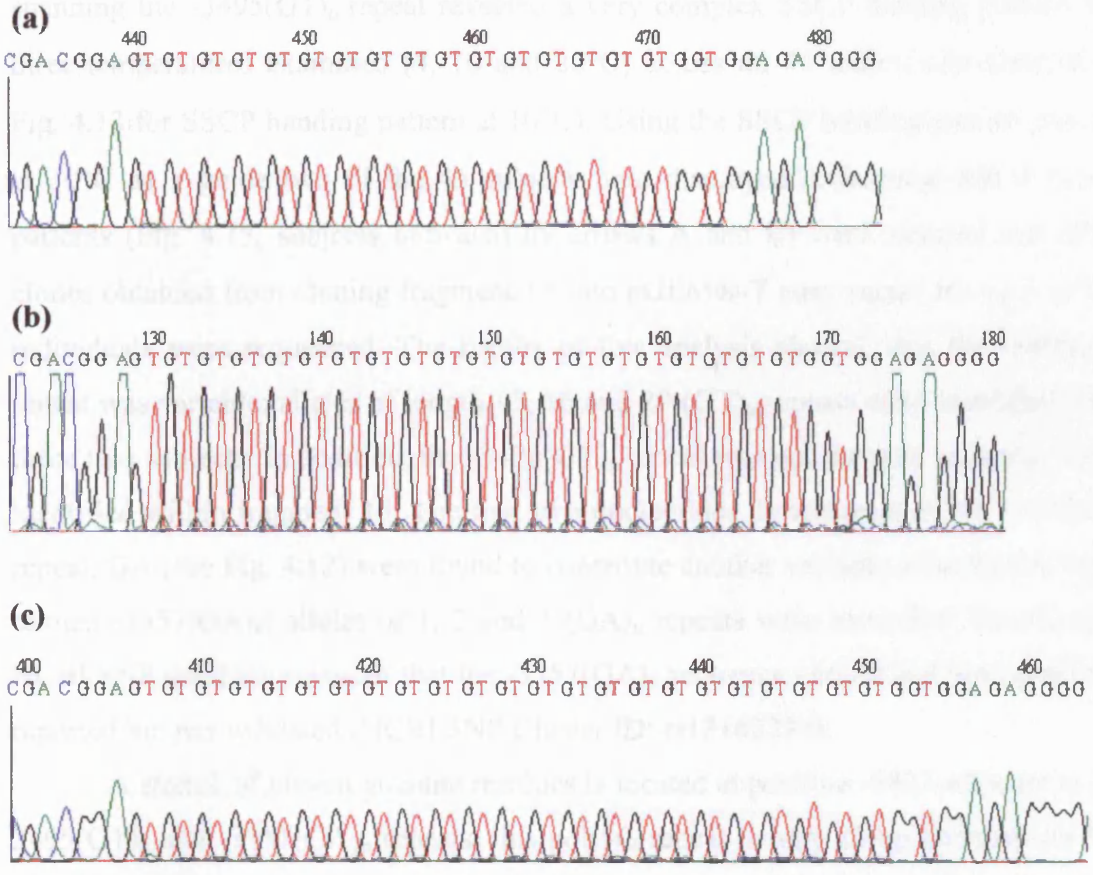


Figure 4.11 Sequencing results for fragment 5 spanning the -556$(GT)_n$ repeat in the Brn-3c 5'-flanking region. Sequencing several clones of an individuals DNA within pGEM®-T easy vector for five subjects that exhibited differential SSCP banding patterns revealed that the -566$(GT)_n$ repeat varied in length from 16 to 23 dincucleotide repeats. Sequencing chromatograms representative of 16 **(a)**, 20 **(b)** and 23 **(c)** repeats are shown. Numbers above each sequencing chromatogram represent the automatic base numbering from the start of readable sequence using BioEdit Sequence Alignment Editor software.

4.4.2 Identification of -3495$(GT)_n$, -3457$(GA)_n$ and -3432poly-G polymorphism.

Having established that the -566$(GT)_n$ repeat is a multi-allelic sequence variant this prompted an examination of a wider region of the published Brn-3c 5'-flanking sequence for the presence of additional simple sequence repeats to see if these too may be polymorphic. Another $(GT)_n$ dinucleotide repeat was located at position -3495 with 19 GT repeats in the reference sequence (Ensembl transcript ID: ENST00000230732; see Appendix B and Fig. 4.12). Therefore, PCR-SSCP analysis was performed using primers 3cP11-S: 5' GCATAAATGCCCACATAGTCC 3' and 3cP11-AS: 5' CCACCACCATGGCATTAAA 3' designed to flank the -3495$(GT)_n$ repeat (fragment 15, Fig. 4.2, page 142) in order to assess whether this region of the Brn-3c promoter (from -3431 to -3561bp) was variable (see Fig. 4.12). PCR-SSCP analysis of the region spanning the -3495$(GT)_n$ repeat revealed a very complex SSCP banding pattern at all three temperatures examined (4, 10 and 20°C) across all 45 individuals scanned (see Fig. 4.13 for SSCP banding pattern at 10°C). Using the SSCP banding pattern produced at 10°C as a guide two of the 45 subjects that displayed differential SSCP banding patterns (Fig. 4.13, subjects indicated by arrows A and B) were selected and several clones obtained from cloning fragment 15 into pGEM®-T easy vector for each of these individuals were sequenced. The results of this analysis showed that the -3495$(GT)_n$ repeat was variable; alleles of length 15, 16 and 20 $(GT)_n$ repeats were identified within these two subjects. In addition, the -3495$(GT)_n$ repeat was not the only sequence variant identified within fragment 15. The first two nucleotides downstream of the -3495$(GT)_n$ repeat, GA (see Fig. 4.12) were found to constitute another variable dinucleotide repeat, termed -3457$(GA)_n$; alleles of 1, 2 and 3 $(GA)_n$ repeats were identified. Searching the NCBI SNP database revealed that the -3457$(GA)_n$ sequence variant had been previously reported but not validated (NCBI SNP Cluster ID: rs13162234).

A stretch of eleven guanine residues is located at position -3432 adjacent to the -3495$(GT)_n$ and -3457$(GA)_n$ repeats; this poly-G repeat is only 23 bp downstream from the -3457$(GA)_n$ repeat (see Fig. 4.12). Mononucleotide repeats as with dinucleotide repeats are abundant in the human genome and are a common source of sequence variation often exhibiting multi-allelic variation (Cohen et al, 2004). Given the close proximity of the -3432 poly-G repeat to the -3457$(GA)_n$ and -3495$(GT)_n$ repeats it was decided to investigate this highly repetitive region of the Brn-3c 5'-flanking region in greater detail. Primers 3cP12-S: 5' GAGCTCCTGAAGCAGTGTCT 3' and 3cP12-AS: 5' GCAATTTGTAACAGCCTCTATGC 3' were designed to amplify a larger

GGAGGATATGGTGCTTTGTGCCCTGACTTCACCTTACATGTGGCAAATCTATGTCCCTACTAGACTCTGA -3742 to
CCTCCTATACCACGAAACACGGGACTGAAGTGGAATGTACACCGTTTAGATACAGGGATGATCTGAGACT -3673

GCTCCTGAAGCAGTGTCTTCTGTATCTTCAAGGCCTTCCACAAGACTCTGCTGTTAGTAGGTATTTGGAA -3672 to
CGAGGACTTCGTCACAGAAGACATAGAAGTTCCGGAAGGTGTTCTGAGACGACAATCATCCATAAACCTT -3603

ATGCTTATCGAATGGGGAAAATTCTTCCTTCAATGCCACTTGCATAAATGCCCACATAGTCCTTACATGG -3602 to
TACGAATAGCTTACCCCTTTTAAGAAGGAAGTTACGGTGAACGTATTTACGGGTGTATCAGGAATGTACC -3533

ACCTTCACATATTTTGTTTGGCCCATAAATGGTATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG -3532 to
TGGAAGTGTATAAAACAAACCGGGTATTTACCATACTCACACACACACACACACACACACACACACAC -3463

TGTGTGATTGTAATTTAATGCCATGGTGGTGGGGGGGGGGGTGGGTAAGTGCTCTCTAGTTTACCACAGT -3462 to
ACACACTAACATTAAATTACGGTACCACCACCCCCCCCCCACCCATTCACGAGAGATCAAATGGTGTCA -3393

TTCAATCTTTCCTTTTATCCACCCAGATAAAGTTGCTCAATTTTATTTTACTAGTTTTGCACCTTATAGA -3392 to
AAGTTAGAAAGGAAAATAGGTGGGTCTATTTCAACGAGTTAAAATAAAATGATCAAAACGTGGAATATCT -3323

ATTTAAGTTTACAAATTCTGGAAAGCATAGAGGCTGTTACAAATTGCCGTCTAAGGCTGTCCTTCCAATT -3322 to
TAAATTCAAATGTTTAAGACCTTTCGTATCTCCGACAATGTTTAACGGCAGATTCCGACAGGAAGGTTAA -3253

GCTTTTTCTATGGTCTTTTCCCTCTTTTCTTCTCCAGTCTCCAGTCTGTTATCTCCAAGGCCTTCTCAAA -3252 to
CGAAAAAGATACCAGAAAAGGGAGAAAAGAAGAGGTCAGAGGTCAGACAATAGAGGTTCCGGAAGAGTTT -3183

Figure 4.12 Section of the Brn-3c 5'-flanking region with loci for the -3495(GT)$_n$ repeat, -3457(GA)$_n$ repeat and -3432 poly-G element highlighted. The -3495(GT)$_n$ repeat and -3432 poly-G element are shown in black bold, the -3457(GA)$_n$ repeat in red bold. Sequence of primers 3cP11-S and 3cP11-AS used in the original PCR-SSCP screen are highlighted in blue with direction of primer extension defined by the blue arrows. Sequence of primers 3cP12-S and 3cP12-AS designed to amplify a larger fragment of the Brn-3c 5'-flanking region incorporating the -3432 poly-G element are highlighted in green with direction of primer extension defined by the green arrows.

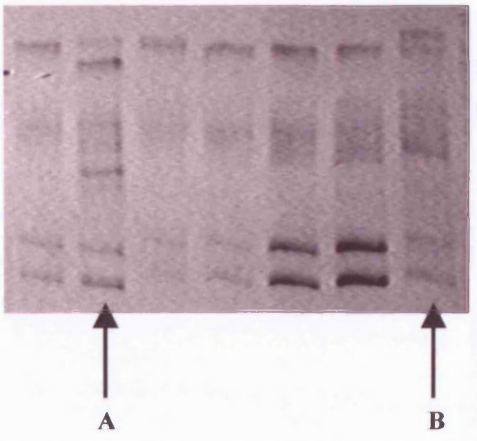


Figure 4.13 SSCP results for fragment 15 spanning the -3495(GT)$_n$ repeat in the Brn-3c 5'-flanking region. Section of the SSCP gel electrophoresed at 10°C to show the complex differential banding pattern; two subjects with distinct banding patterns are indicated by arrows A and B as an example. Primer sequences for fragment 15 are shown in Table 3.1 Method section 3.2.18.

fragment (from -3276 to -3674bp) that in addition to the -3495(GT)$_n$ and -3457(GA)$_n$ repeats also contained the poly-G repeat at -3432 (Fig. 4.12; see also Method section 3.2.2.4). Genomic DNA from six subjects that originally displayed a differential SSCP banding pattern for the -3431 to -3561bp region encompassing the -3495(GT)$_n$ and -3457(GA)$_n$ repeats (Fig. 4.12) was amplified by PCR using primers 3cP12-S and 3cP12-

AS and several clones obtained from cloning this larger fragment (from -3276 to -3674bp) into pGEM®-T easy vector for each of these six different individuals were sequenced. These results, coupled with subsequent genotyping analysis for a preliminary association study (see Chapter 8) showed that the -3495(GT) repeat ranged in length from 15 to 21 dinucleotides, with an additional allele of 24 dinucleotides [-*3495(GT)*$_{15\text{-}21,24}$] and the -3457(GA) repeat from 1 to 3 dinucleotides [-*3457(GA)*$_{1\text{-}3}$]. In addition the -3432poly-G repeat located just 23bp downstream of the -3457(GA)$_n$ repeat (see Fig. 4.12) was found to constitute a highly complex and common sequence variant exhibiting multiple variations in length together with single nucleotide substitutions within the poly-G repeat (see Fig. 4.14). Originally, five different alleles were identified for the -3432poly-G repeat as part of the initial PCR-SSCP screen (see Table 4.3; poly-G alleles that are highlighted in blue) and eventually a total of nine different alleles were identified based on the subsequent genotyping analysis for a preliminary association study (see Table 4.3). This sequence variant was termed the *-3432poly-G polymorphism*. Four of the alleles identified are simple variations in mononucleotide length termed (G)$_x$ where x was found to vary between 10 and 13 guanine repeats. The remaining five alleles all display single nucleotide substitutions termed SNPGX in addition to varying in length (see Table 4.3)

Table 4.3: The polymorphic alleles identified at position -3432 within the Brn-3c 5'-flanking region.

Allele	DNA sequence	Abbreviation	Length of allele
(G)$_{10}$	5' TGGTGGT**GGGGGGGGGG**TGGGTA 3'	(G)$_{10}$	10
(G)$_{11}$	5' TGGTGGT**GGGGGGGGGGG**TGGGTA 3'	(G)$_{11}$	11
(G)$_{12}$	5' TGGTGGT**GGGGGGGGGGGG**TGGGTA 3'	(G)$_{12}$	12
(G)$_{13}$	5' TGGTGGT**GGGGGGGGGGGGG**TGGGTA 3'	(G)$_{13}$	13
SNPG1	5' TGGTGGT**GGGGGGGGCG**TGGGTA 3'	(G)$_8$CG	10
SNPG2	5' TGGTGGT**GGCGGGGGGGGG**TGGGTA 3'	GGC(G)$_9$	12
SNPG4	5' TGGTGGT**GGTGGGGGGGG**TGGGTA 3'	GGT(G)$_8$	11
SNPG5	5' TGGTGGT**GGGGCGGGGGG**TGGGTA 3'	(G)$_4$C(G)$_6$	11
SNPG6	5' TGGTGGT**GGCGGGGGGGGGG**TGGGTA 3'	GGC(G)$_{10}$	13

Alleles identified are based on PCR-SSCP analysis in 45 random individuals predominantly of Caucasian ethnicity in addition to 142 patients characterised with late onset hearing loss and 145 general population samples genotyped for a preliminary association study. All subjects used in the preliminary association study are Caucasian (see Chapter 8 for details of the association study). Alleles highlighted in blue were initially identified as part of the PCR-SSCP screen. Alleles identified either vary in poly-guanine repeat length or exhibit single nucleotide substitutions (shown in red bold) within the poly-guanine repeat.

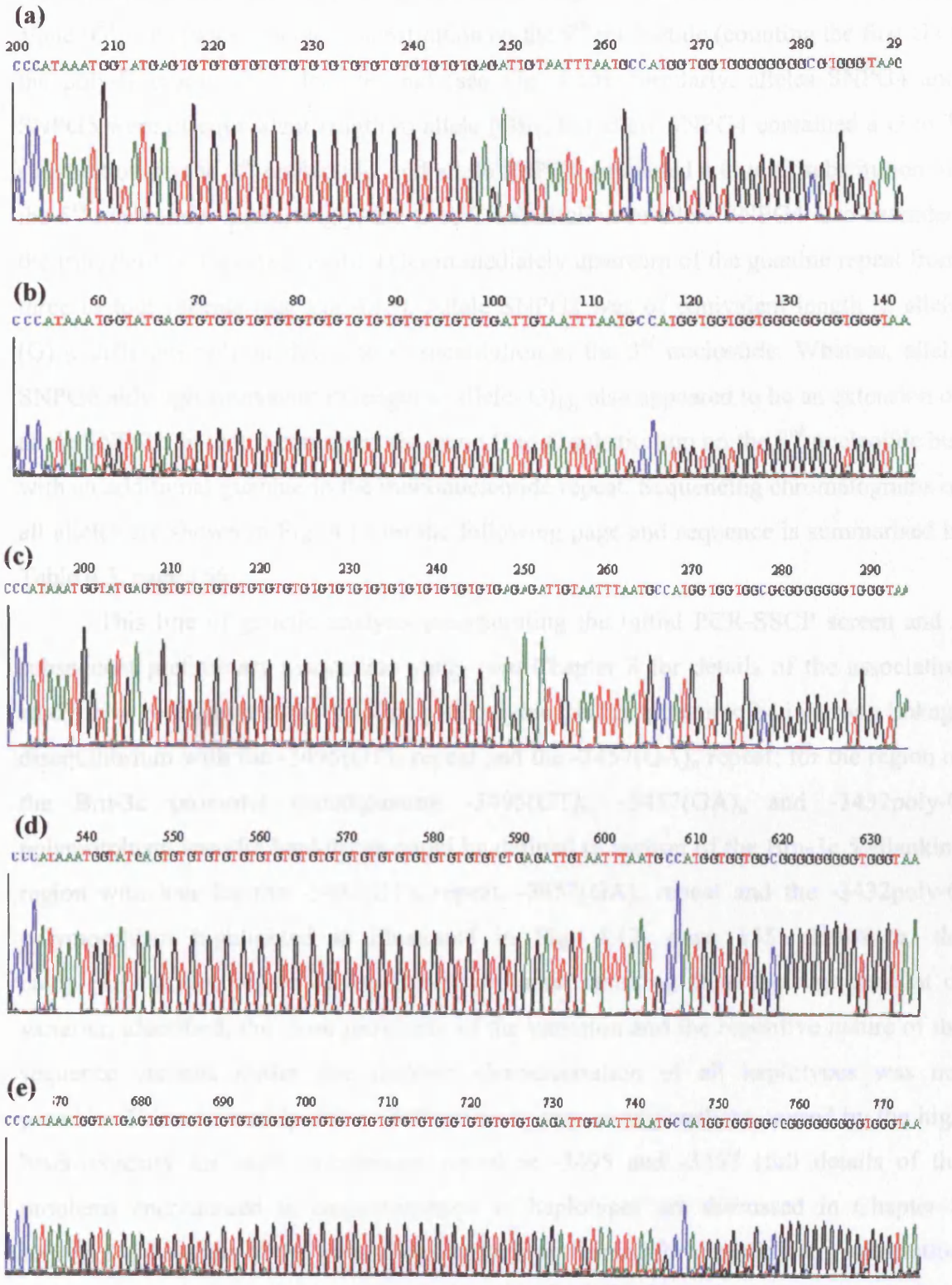


Figure 4.14 Representative sequencing chromatograms to illustrate three multi-allelic sequence variants in the Brn-3c 5'-flanking region: -3495(GT)$_n$, -3457(GA)$_n$ and -3432poly-G polymorphism. Sequencing chromatograms shown were obtained from sequencing cloned PCR products in pGEM®-T easy vector for different individuals as part of an initial PCR-SSCP screen and a subsequent preliminary association study. The sense strand of sequence is shown. Numbers above each sequencing chromatogram represent the automatic base numbering from the start of readable sequence using BioEdit Sequence Alignment Editor software.

(a) -3495(GT)$_{15}$: -3457(GA)$_2$: -3432(G)$_8$CG, **(b)** -3495(GT)$_{16}$: -3457(GA)$_1$: -3432GGT(G)$_8$,
(c) -3495(GT)$_{20}$: -3457(GA)$_3$: -3432GGC(G)$_9$, **(d)** -3495(GT)$_{21}$: -3457(GA)$_2$: -3432GGC(G)$_9$,
(e) -3495(GT)$_{24}$: -3457(GA)$_2$: -3432GGC(G)$_{10}$.

For example allele SNPG1 although the same length as the shortest allele identified, allele $(G)_{10}$, contained a G to C substitution on the 9th nucleotide (counting the first G of the poly-G repeat as nucleotide one) (see Fig. 4.15). Similarly, alleles SNPG4 and SNPG5 were of equivalent length to allele $(G)_{11}$, but allele SNPG4 contained a G to T substitution on the 3rd nucleotide, and allele SNPG5 contained a G to C substitution on the 5th nucleotide. Interestingly, the G to T substitution in allele SNPG4 also extended the trinucleotide repeating motif TGG immediately upstream of the guanine repeat from three to four repeats (see Fig 4.15). Allele SNPG2 was of equivalent length to allele $(G)_{12}$, differing only in the G to C substitution at the 3rd nucleotide. Whereas, allele SNPG6 although equivalent in length to allele $(G)_{13}$, also appeared to be an extension of allele SNPG2, in that it contained the same G to C substitution on the 3rd nucleotide but with an additional guanine in the mononucleotide repeat. Sequencing chromatograms of all alleles are shown in Fig. 4.15 on the following page and sequence is summarised in Table 4.3, page 156.

This line of genetic analysis incorporating the initial PCR-SSCP screen and a subsequent preliminary association study (see Chapter 8 for details of the association study) also revealed that the -3432poly-G polymorphism appears to be in strong linkage disequilibrium with the -3495$(GT)_n$ repeat and the -3457$(GA)_n$ repeat; for the region of the Brn-3c promoter encompassing -3495$(GT)_n$, -3457$(GA)_n$ and -3432poly-G polymorphism specific haplotypes could be defined (a section of the Brn-3c 5'-flanking region with loci for the -3495$(GT)_n$ repeat, -3457$(GA)_n$ repeat and the -3432poly-G polymorphism highlighted is illustrated in Fig. 4.12, page 155). However, the complexity of this region of the Brn-3c promoter resulting from the high amount of variation identified, the close proximity of the variation and the repetitive nature of the sequence variants meant that detailed characterisation of all haplotypes was not possible. This was mainly due to difficulties in sequencing analysis caused by the high heterozygosity for each dinucleotide repeat at -3495 and -3457 (full details of the problems encountered in characterisation of haplotypes are discussed in Chapter 8 section 8.4.4 as part of the genotyping analysis performed for a preliminary association study). However, some haplotypes could be clearly defined based on genetic analysis of subjects homozygous for the -3432poly-G polymorphism. For example, it was evident that if allele SNPG1 was present at position -3432 the native 5' haplotype detected always consisted of -3457$(GA)_2$: -3495$(GT)_{15}$ (Fig. 4.14a, page 157). Similarly, if allele SNPG2 was present at position -3432 the common 5' haplotype detected was always -3457$(GA)_2$: -3495$(GT)_{20\text{ or }21}$ (Fig. 4.14d, page 157). Haplotypes that it was

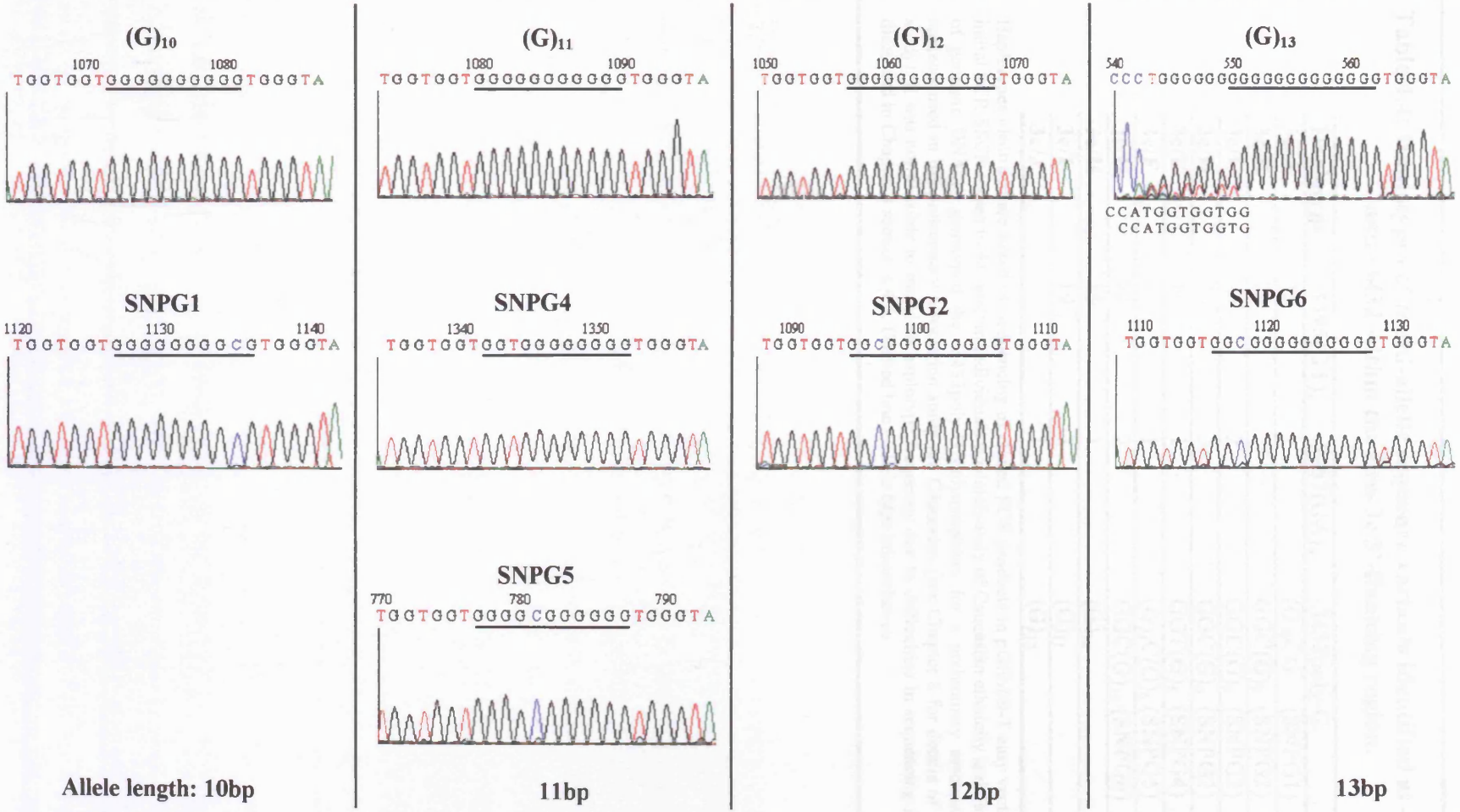


Figure 4.15: Sequencing chromatograms to illustrate the multi-allelic -3432poly-G polymorphism in the Brn-3c 5'-flanking region. Alleles are grouped according to length of the poly-G repeat, which is underlined for each allele. Chromatograms shown represent sequencing of cloned PCR products except for allele $(G)_{13}$ the chromatogram shown represents direct sequencing of genomic DNA from an individual heterozygous for allele $(G)_{13}$: genotype $(G)_{13}/(G)_{12}$. The correct base calls for this sequence are shown below the chromatogram for each allele. In all cases sequencing was performed using primer 3cP12-AS but for convenience the sense strand is shown. Numbers above each sequencing chromatogram represent the automatic base numbering from the start of readable sequence using BioEdit Sequence Alignment Editor software.

possible to decipher are summarised in Table 4.4 and discussed in greater detail in Chapter 8 section 8.4.4.

Table 4.4: Haplotypes of multi-allelic sequence variants identified at -3495, -3457 and -3432 within the Brn-3c 5'-flanking region.

Haplotype	-3495$(GT)_n$	-3457$(GA)_n$	-3432poly-G
3c/A	15	2	$(G)_8CG$ (SNPG1)
3c/B	20	2	$GGC(G)_9$ (SNPG2)
3c/C	21	2	$GGC(G)_9$ (SNPG2)
3c/D	20	3	$GGC(G)_9$ (SNPG2)
3c/E	16	1	$GGT(G)_8$ (SNPG4)
3c/F	-	1	$(G)_4C(G)_6$ (SNPG5)
3c/G	24	2	$GGC(G)_{10}$ (SNPG6)
3c/H	19	1	$(G)_{10}$
3c/I	19	1	$(G)_{11}$
3c/J	-	1	$(G)_{12}$

Haplotypes identified are based on sequencing of cloned PCR products in pGEM®-T easy vector as part of an initial PCR-SSCP screen in 45 random individuals predominantly of Caucasian ethnicity and direct sequencing of genomic DNA to genotyped the -3432poly-G polymorphism for a preliminary association study. All subjects used in the preliminary association study are Caucasian (see Chapter 8 for details of the association study). It was not possible to establish haplotype frequency due to difficulties in sequencing analysis; this is discussed in Chapter 8 section 8.4.5. Dashed line – allele type inconclusive.

4.5 General discussion.

To assess the extent of genetic variation at the Brn-3c locus and to search for common sequence variants mutation scanning by PCR-SSCP analysis was performed on DNA isolated from 45 random individuals. A total of 15 PCR fragments were examined: 6 covering the coding region and intron of the gene, 8 spanning 1.5Kb of the 5'-flanking region, plus one fragment designed to include the dinucleotide $(GT)_n$ repeat at -3495 within the distal 5'-flanking region. Using this approach five novel sequence variants were identified at the Brn-3c locus and the presence of two previously reported but non-validated sequence variants were confirmed. Thus, confirming the suitability of the PCR-SSCP screen for detection of sequence variation. Genetic variation at the Brn-3c locus is summarised in Fig. 4.16 and was deposited in the NCBI SNP database (NCBI SNP Cluster ID for each sequence variant is given in the legend to Fig. 4.16).

Of the novel sequence variants identified at the Brn-3c locus, two are simple bi-allelic single nucleotide substitutions: +90C>T within exon 1 and -386C>A within the 5'-flanking region of the gene. The remaining novel variants are multi-allelic sequence variants: two variable dinucleotide $(GT)_n$ repeats -566$(GT)_{17\text{-}23}$ and -3495$(GT)_{15\text{-}21,\ 24}$ both within the 5'-flanking region of the gene and a highly complex polymorphism, a guanine repeat at position -3432 that in addition to displaying variations in length also exhibits single base substitutions within the repeat (-3432poly-G polymorphism). A third dinucleotide repeat consisting of 5 CT repeats in the intron of Brn-3c (for reference sequence see Appendix B) was not found to be variable based on the results of this PCR-SSCP screen. Analysis of the Brn-3c locus by PCR-SSCP analysis followed by sequencing analysis also confirmed two previously reported but non-validated sequence variants: a single nucleotide substitution at -1391, A>C (NCBI SNP Cluster ID: rs1368402) and a variable dinucleotide repeat at -3457, $(GA)_{1\text{-}3}$ (NCBI SNP Cluster ID: rs13162234); both are within the 5'-flanking region of the Brn-3c gene.

The aim of this project is to assess whether common sequence variants in the Brn-3c gene are a risk factor for late onset hearing loss exhibited by a large proportion of the ageing population by using only those sequence variants where there is experimental evidence to suggest they are functional in subsequent case-control association analysis. The results of this PCR-SSCP screen coupled with sequencing analysis suggest that five sequence variants identified at the Brn-3c locus are common in the general population: -3495$(GT)_{15\text{-}21,24}$, -3457$(GA)_{1\text{-}3}$, -3432poly-G polymorphism, -1391A >C and -566$(GT)_{17\text{-}23}$; all are within the 5'-flanking region of the Brn-3c gene.

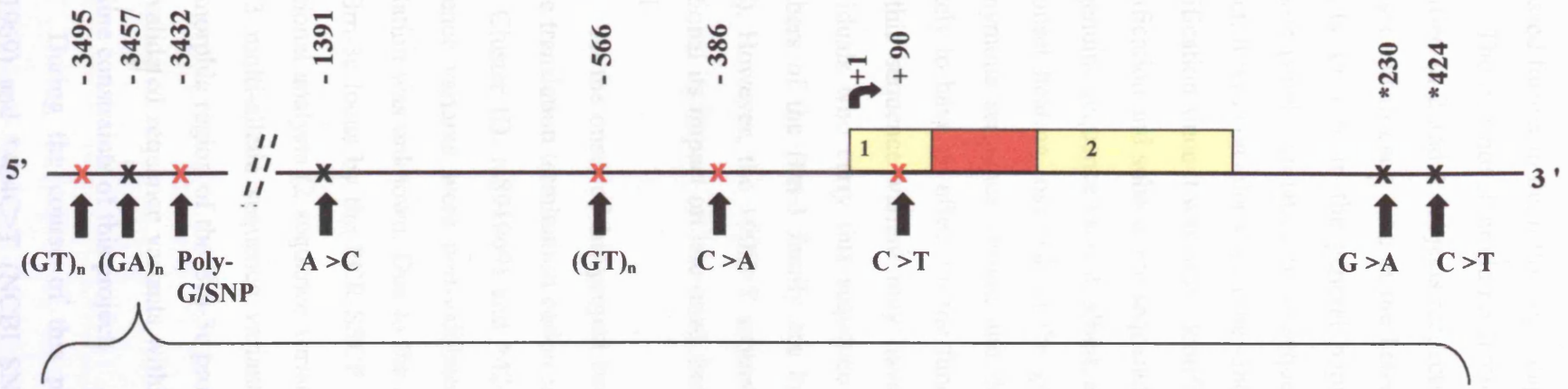


Figure 4.16 A schematic illustration of genetic variation identified at the Brn-3c locus.
Variants are numbered with respect to the first A of the ATG translation start site as +1 (indicated by a curved black arrow). Novel variants are highlighted in red. Variants reported but non-validated at the onset of this project are highlighted in black; of these -1391A >C and -3457$(GA)_n$ were confirmed as part of the PCR-SSCP screen and *230G>A and *424C>T were confirmed independently (the nature of the sequence variation is indicated below). Exons 1 and 2 are numbered and shown as pale yellow boxes, the intron as a red shaded box (not to scale). For simplicity the region of the Brn-3c promoter between -1391 and -3495 has been omitted as indicated by the two hatched lines. The sequence containing the three multi-allelic sequence variants at -3.4Kb is shown: the -3495$(GT)_n$ repeat and -3432poly-G polymorphism are highlighted in bold, the -3457$(GA)_n$ repeat in red bold, the allele shown is: -3495$(GT)_{19}$: -3457$(GA)_1$: -3432$(G)_{11}$ which forms haplotype 3c/I (see Table 4.4). NCBI SNP Cluster ID: -3495$(GT)_n$ (rs28994882), -3457$(GA)_n$ (rs28994881), -3432poly-G polymorphism (rs29001168), -1391A>C (rs1368402), -566$(GT)_n$ (rs28987086), -386C>A (rs28994880), +90C>T (rs28994879), *230G>A (rs891969) , *424C>T (rs891970).

Therefore, in accord with the aim of this project (see section 2.0) these sequence variants were considered candidates for subsequent functional analysis and are discussed further in the following Chapters, 5 to 7.

The two novel single nucleotide substitutions +90C>T and -386C>A were only identified in 2 and 1 subjects respectively, of the 45 subjects examined by PCR-SSCP analysis and in each case in the heterozygous state. This suggests that these sequence variants are rare in the general population and since the purpose of this project is *common* genetic variation these sequence variants were not pursued further within this project. It was considered that the -386C>A substitution could be an artefact of the PCR amplification since it was only identified in one individual. However, repeating the PCR amplification and subsequent sequencing analysis ruled out this possibility confirming it is a genuine sequence variant, albeit, a rare variant and unlikely to have much impact on late onset hearing loss risk in the general population. The +90C>T substitution is a synonymous sequence variant and does not change an amino-acid and therefore, is unlikely to have an affect on the function of Brn-3c. Although, it cannot be ruled out that this sequence variant may have an effect on alternative splicing of Brn-3c in individuals who carry this sequence variant; both Brn-3a and Brn-3b the additional members of the Brn-3 family are known to undergo alternative splicing (Liu et al, 1996). However, the +90C>T sequence variant appears rare and even if found to be functional its impact on late onset hearing loss risk in the general population would be small.

At the onset of this project two single nucleotide substitutions were reported 3' of the translation termination codon in the Brn-3c 3'-flanking region: *230G>A (NCBI SNP Cluster ID: rs891969) and *424C>T (NCBI SNP Cluster ID: rs891970). These sequence variants were non-validated and their significance with respect to Brn-3c regulation was unknown. Due to the amount of common genetic variation identified at the Brn-3c locus by the PCR-SSCP screen that were good candidates for subsequent functional analysis (2 sequence variants within 1.5Kb of the Brn-3c 5'-flanking region and 3 multi-allelic sequence variants in close proximity at a highly repetitive and polymorphic region of the Brn-3c promoter at -3.4Kb) it was decided not to pursue the non-validated sequence variants within the 3'-flanking region of Brn-3c further within the time constraints of this project.

During the course of this project both *230G>A (NCBI SNP Cluster ID: rs891969) and *424C>T (NCBI SNP Cluster ID: rs891970) within the 3'-flanking region of Brn-3c have been validated independently with a rare allele frequency of 0.27

and 0.283 respectively, in samples of Caucasian ancestry (see NCBI dbSNP Human Build 126; http://www.ncbi.nlm.nih.gov/project/SNP/). Therefore, given that these sequence variants are common it should be considered that they may be a risk factor for late onset hearing loss in the general population. It is possible they may act as downstream modulators of Brn-3c promoter activity and cause inter-individual differences in Brn-3c expression. Indeed, recently a nuclear factor of activated T cells CCAAT/enhancer-binding protein (NFAT.C/EPB) composite element has been identified in the Brn-3c 3'flanking region (Yang et al, 2003). A search of the NCBI SNP database revealed that no previously reported single nucleotide substitutions intercepted this element but the *230G>A substitution is only 5bp downstream. The NFAT.C/EPB composite element was originally identified as a novel regulatory enhancer in the immediate upstream region of the peroxisome proliferator-activated receptor-γ2 promoter and it was through subsequent promoter analysis that highly similar derivatives of the NFAT.C/EPB composite element were found in the regulatory regions of several other genes including Brn-3c (Yang et al, 2003). Electromobility shift analysis (EMSA) using nuclear extracts prepared from COS cells transfected with NFATc4 and / or C/EBPα expression constructs suggests that both NFAT and C/EPB bind to the NFAT.C/EPB composite element in the Brn-3c 3'-flanking region and reporter gene analysis in BHK cells suggests that this element is functional (Yang et al, 2003). However, the significance of this element on Brn-3c expression in sensory hair cells is not known. Nevertheless, these findings do raise the possibility that the NFAT.C/EPB composite element in the Brn-3c 3'-flanking region could be functional in sensory hair cells and important for regulation of the Brn-3c gene. Moreover, the finding that the common *230G>A substitution is only 5bp downstream raises the possibility that this sequence variant could modify protein binding to the NFAT.C/EPB composite element and thus have an affect on regulation of Brn-3c; protein binding to cis-acting elements is influenced by adjacent flanking sequence (Dawson et al, 1996a). It should also be considered that *230G>A and *424C>T may be important in post-transcriptional control of Brn-3c regulation for example, by affecting mRNA stability. However, the 3' end of the mRNA transcript for Brn-3c has not be defined in humans and it is not known whether *230G>A and *424C>T are in the 3'-UTR of Brn-3c. If *230G>A and *424C>T are found to manifest in the 3'-UTR of Brn-3c it is possible they may affect stability of the Brn-3c mRNA transcript by modifying protein binding to important cis-acting elements within the 3'UTR of Brn-3c. However, this can only be speculated upon at present.

4.6 Conclusion.

In conclusion, the data presented here within this thesis is the first comprehensive report of genetic variation in the Brn-3c gene, a hair cell pro-survival factor and a good candidate gene for late onset hearing loss. Mutation scanning of the Brn-3c gene by PCR-SSCP analysis has identified five novel sequence variants at the Brn-3c locus and confirmed the presence of two that were previously reported but non-validated (see Fig. 4.16, page 162). Of the variants identified and confirmed, five appear common in the general population and are candidates for functional analysis. Interestingly, all common variants are within the 5'-flanking region of the Brn-3c gene raising the possibility they could modulate regulation of Brn-3c leading to inter-individual variations in Brn-3c expression. Common sequence variants that have a functional affect on regulation of Brn-3c could be a risk factor for late onset hearing loss; this line of analysis is discussed in the following chapters.

5.0 Functional Characterisation Of The -3432poly-G Polymorphism At The Brn-3c Locus.

5.1 Introduction.

Mutation scanning of the Brn-3c gene by PCR-SSCP analysis identified five novel sequence variants at the Brn-3c locus and confirmed the presence of two that were previously reported but non-validated (see Fig. 4.16, section 4.5, page 162). One novel sequence variant was identified within exon one of the Brn-3c gene, +90C>T, but this sequence variant appeared rare in the general population and was found to constitute a synonymous variant with both alleles encoding for the amino acid alanine and so was not pursued further. The remaining six sequence variants were identified within the 5'-flanking region of the Brn-3c gene and of these five: -3495$(GT)_{15\text{-}21,\ 24}$, -3457$(GA)_{1\text{-}3}$, -3432poly-G polymorphism, -1391A>C and -566$(GT)_{17\text{-}23}$ appear common in the general population (based on PCR-SSCP analysis on 45 individuals; see section 4.5).

Common sequence variants are the basis of the CD:CV hypothesis that proposes they underlie susceptibility to common disease (for reviews see Smith and Lusis, 2002; Wright et al, 2003). It is possible than any or a combination of common sequence variants identified in the Brn-3c 5'-flanking region could be a risk factor for late onset hearing loss exhibited by a large proportion of the ageing population; they could modulate regulation of the Brn-3c gene leading to inter-individual variations in Brn-3c expression. Indeed, sequence variants that manifest within the promoter of a gene can have an adverse affect on expression of the downstream gene by disrupting existing or creating novel cis-acting elements and hence, modify the binding of trans-acting factors. Certainly, there is increasing evidence that functional sequence variants involving a single nucleotide change in the DNA sequence within gene regulatory regions are involved in susceptibility to common, complex disease. For example, a C to T substitution at -1562 in the matrix metalloproteinase-9 (MMP-9) promoter that is associated with severity of coronary atherosclerosis modifies the binding affinity of a nuclear protein at -1562 in the MMP-9 promoter such that greater promoter activity is observed for the T allele compared to the C allele when tested using a reporter gene assay (Zhang et al, 1999). Similarly, a G to A substitution at -112 in the uteroglobin-related protein 1 (UGRP1) promoter that modifies nuclear protein binding is associated

with an increased risk of asthma (Niimi et al, 2002). The G allele has a greater affinity for a nuclear protein compared to the A allele and this is consistent with a 24% reduction in UGRP1 promoter activity observed for the A allele compared to the G allele when examined by reporter gene analysis. Likewise, a G to A substitution at -455 in the β-fibrinogen promoter, which modifies nuclear protein binding, has a functional affect on interleukine-6 stimulated activity of the β-fibrinogen promoter and the -455A allele has been consistently associated with elevated plasma fibrinogen levels and atherothrombotic disease (for review see Green, 2001). Finally, genetic variation in the PAI-1 gene promoter has for some time been known to affect regulation of the PAI-1 gene; the 4G/5G insertion / deletion at -675 is thought to modify binding of a transcriptional repressor (Dawson et al, 1993) and this sequence variant has been associated with susceptibility to cardiovascular disease (for reviews see Nordt et al, 2001; Dellas and Loskutoff, 2005).

Therefore, to determine whether any of the common sequence variants identified within the 5'-flanking region of the Brn-3c gene underlie susceptibility to late onset hearing loss, in the first instance, functional analysis will be performed to assess whether they have an affect on regulation of the Brn-3c gene. Measures will be taken to assess whether any of these common sequence variants modify the binding of transcription factors and whether this translates into functional differences in Brn-3c promoter activity. Functional sequence variants will then be used in subsequent case-control association analysis to assess whether they are a genetic risk factor for late onset hearing loss.

5.1.1 The electrophoretic mobility shift assay.

To assess whether common sequence variants identified within the promoter of the Brn-3c gene modulate the binding of a transcription factor or factors, protein-DNA interactions will be investigated by use of the electrophoretic mobility shift assay (EMSA). Originally devised by Fried and Crothers (1981) and Garner and Revzin (1981) the EMSA assay provides a rapid and simple means for detecting proteins binding to specific DNA sequences. EMSA analysis involves the incubation of radiolabelled DNA fragments as probes with whole cell or nuclear protein extracts and is based on the principle that protein-DNA complexes are resolved more slowly by electrophoresis through non-denaturing polyacrylamide gels than free DNA probe alone. Oligonucleotide probes can be devised that contain either the wild-type or variant

allele for any given sequence variant and the effect of the sequence variant on protein binding can be assessed and visualised by autoradiography. If a sequence variant induces differential protein binding steps can then be taken to assess the specificity of the interaction and the relative affinity of each allele for the protein(s) involved. This analysis can be followed by measures to identify the protein(s) binding including competition analysis with known transcription factor consensus sequences and / or supershift analysis with specific antibodies.

5.1.2 The reporter gene assay.

EMSA analysis is very useful for determining sequence-specific protein binding and when used in combination with supershift analysis identification of the proteins involved. However, determination of sequence-specific protein binding does not mean that the cis-acting element in question is functional. Therefore, if a sequence variant modulates the binding of a specific transcription factor this should always be followed-up with functional analysis to assess whether the differential protein binding actually has an affect on regulation or expression of the gene, or both. The examination of the effect of sequence variation on gene regulation can be achieved *in-vitro* by use of reporter gene assays. For a bi-allelic sequence variant in the promoter of a gene this generally involves sub-cloning the promoter sequence under test upstream of a reporter gene carried in an expression vector and making two constructs,one containing the wild-type allele and the other the variant allele. These reporter gene constructs are then transfected into a suitable cell line and assayed for reporter gene expression; in principle the amount of reporter gene expression being dependent on the regulatory sequence and hence, the allele cloned upstream of the reporter gene. Although, one common drawback of the reporter gene assay is that differences in levels of reporter gene activity between two constructs can arise not because of functional differences in activity of the promoter sequence under test but because of differences in transfection efficiency between the constructs. However, this problem can be minimised by co-transfection of an internal control reporter gene construct to normalise for differences in transfection efficiency within and between experiments. There are many different types of commercially available reporter gene systems to study gene regulation. One of the most common reporter genes used in mammalian cells is the firefly (*Photinus pyralis*) luciferase enzyme (de Wet et al, 1987). Reporter gene systems incorporating luciferase enzymes exploit the fact that these enzymes catalyse bioluminescent reactions and

produce photons of light, which can be quantified with a luminometer. Thus, producing rapid and sensitive reporter gene assays. For functional analysis of common sequence variants identified within the Brn-3c promoter the dual-luciferase® reporter assay system (Promega) was chosen. This dual system has an added advantage in that it allows for the sequential assay of both firefly luciferase (the experimental reporter gene joined to the promoter sequence under test) and sea pansy (*Renilla reniformis*) luciferase (the internal control reporter gene) within a single sample (Sherf et al, 1996; Wood, 1998). Firefly and *renilla* luciferases are evolutionary diverse and are not alike structurally nor in substrate specificity. Consequently, it is possible to distinguish between their individual bioluminescent reactions in a single sample. By using the dual-luciferase® reporter assay system (Promega), "the luminescence from the firefly luciferase reaction can be quenched while in parallel activating the luminescent reaction of *renilla* luciferase" (Sherf et al, 1996; Wood, 1998). Normalising the activity of firefly luciferase (the experimental reporter) to that of *renilla* luciferase (the internal control reporter) intra- and inter- experimental variation resulting from differences in transfection efficiency or cell viability can be reduced and variability arising from preparation of extracts for assay can be removed. Thereby, allowing for accurate and reliable interpretation of experimental data generated.

It is important to consider that this *in-vitro* methodology incorporating EMSA analysis in combination with the reporter gene assay is limited in that it cannot provide evidence and establish differences in gene expression *in-vivo* (the limitations of these techniques are discussed further in section 5.8.1). However, this *in-vitro* approach does provide insight into the possible functional effect of sequence variation on gene regulation and can shed light on the molecular pathway(s) involved.

5.1.3 The OC-2 cell line used for functional analysis in this project: background.

The OC-2 (organ of Corti-2) cell line derived from inner ear sensory epithelium was used for functional analysis of common sequence variants in the Brn-3c gene and kindly provided by Professor Matthew Holley, University of Sheffield. The OC-2 cell line along with the OC-1 cell line is derived from the H-2K^{b}-*ts*A58 transgenic mouse, Immortomouse (Rivolta et al, 1998). Both OC-1 and OC-2 cell lines are of immense value to auditory research. Hair cells are present in extremely low numbers in the mammalian cochlea (in the region of 16,000 are present per cochlea in humans) and

experimentally they are inaccessible. These factors taken together with the fact that hair cells do not naturally proliferate *in-vitro* have been a major hindrance to progress in auditory research.

The Immortomouse carries a conditional immortalizing transgene, a temperature sensitive mutant of the tumour antigen from the simian virus 40, tsA58, under the control of the mouse major histocompatibility complex (MHC) H-$2K^b$ class 1 promoter. The MHC class 1 promoter directs expression of the immortalizing gene to a broad range of tissues and expression can be stimulated to high levels in almost all cell types by exposure to γ-interferon (γ-IFN) (Jat et al, 1991). Hence, cells derived from Immortomouse and cultured under permissive conditions: 33°C in the presence of γ-IFN, will continue to proliferate. If the culture conditions are changed to the non-permissive conditions by withdrawal of γ-IFN and increasing the temperature to 39°C, the cells will cease proliferation and start to differentiate; the tsA58 gene product is degraded upon change to the higher temperature (Jat et al, 1991). The organ of Corti derived cell lines were isolated by inducing expression of the immortalizing gene in organotypic cultures of developing organ of Corti (by incubation at 33°C in the presence of γ-interferon, γ-IFN) at E13 immediately before hair cells undergo their final mitosis (Rivolta et al, 1998). Characterisation of the OC-1 and OC-2 cell lines under immortalizing and differentiating conditions reveals that each expresses hair cell specific markers (Rivolta et al, 1998). Under immortalising conditions OC-2 cells express Brn-3c, the α9 subunit of the nicotinic acetyl-choline receptor (α9AChR) and the cytoskeletal hair cell components fimbrin, myosin VI and myosin VIIa. Under the same conditions in OC-1 cells, there is little expression of α9AChR and myosin VIIa and Brn-3c mRNA can hardly be detected however under differentiating conditions these hair cell markers are upregulated. Consequently, it is suggested that OC-1 cells are derived from an earlier stage of development compared to OC-2. In particular, it is proposed that OC-1 cells were immortalized at a stage after specification of the hair cell fate but just before these committed hair cells begin to express Brn-3c (Rivolta et al, 1998).

Since their isolation in 1998, both OC-1 and OC-2 cell lines have been used in many aspects of functional auditory research (for review see Rivolta and Holley, 2002). OC-1 and OC-2 cells have been used to identify functional regions of Brn-3c to aid our understanding of how Brn-3c may function in inner ear sensory hair cells, and to examine the mechanism whereby Barhl1, a hair cell survival factor, may regulate gene transcription in hair cells (Sud et al, 2005; see section 1.9.2). In humans, an 8bp deletion

in Brn-3c is responsible for DFNA15 (Vahava et al, 1998), and use of the OC-2 cell line has helped elucidate the effect of the 8bp deletion on the function of the Brn-3c protein (Weiss et al, 2003, Clough et al, 2004; see section 1.8.2). Importantly, use of both these cell lines has helped provide evidence to suggest that Brn-3c directly regulates BDNF and NT-3 transcripts in hair cells (Clough et al, 2004; see section 1.9.2). Since OC-2 cells strongly express Brn-3c under immortalizing conditions it was reasoned that trans-acting factors needed for Brn-3c expression would be present. Therefore, the OC-2 cell line was regarded a suitable medium to characterise common sequence variants identified in the Brn-3c promoter and was used for preparation of nuclear extract for use in EMSA analysis and for performing transient transfection reporter gene assay.

In summary, identification of a common sequence variant or common haplotype in the Brn-3c promoter that is functional would be highly significant as it may be a risk factor for late onset hearing loss; a factor that can be tested by case-control association analysis. It would also pave way for understanding how the Brn-3c gene, an important hair cell pro-survival factor is regulated at the level of transcription – something for which currently very little is known.

5.2 The -3432poly-G polymorphism in the Brn-3c promoter: similarity to a complex poly-G length polymorphism identified in the tissue kallikrein promoter.

The -3432poly-G polymorphism in the Brn-3c promoter is extremely complex exhibiting multiple variations in length coupled with single nucleotide substitutions within the poly-G repeat (see Table 4.3, section 4.4.2, page 156). Given the unique complexity of the -3432poly-G polymorphism a literature search was undertaken to establish whether any similar polymorphic elements had been described. This line of analysis revealed that the -3432poly-G polymorphism is very similar to a poly-G length polymorphism that has been identified at -121 to -131bp in the promoter of the tissue kallikrein (KLK1) gene (Song et al, 1997; Yu et al, 2002). The KLK1 gene encodes a serine proteinase and is the main enzyme responsible for kinin generation in the kidney. Kinins are potent vasodilators and have long been implicated in the pathophysiology of human disease including essential hypertension (for review see Campbell, 2001). Currently, on a mixed ethnic background 15 alleles have been identified for the poly-G length polymorphism in the KLK1 promoter including similar poly-guanine alleles (guanine number ranging from 10 to 13) and alleles that exhibit similar single nucleotide substitutions to those identified in the Brn-3c promoter. A summary of the poly-G alleles identified in the KLK1 promoter is shown in Table 5.1 and for comparison the similar poly-G alleles identified in the Brn-3c promoter are also shown.

Reporter gene assays in the human embryonic kidney 293 cell line have shown that the KLK1 poly-G polymorphism has a functional effect on KLK1 promoter activity with specific poly-G alleles exhibiting significantly lower promoter activities compared to others. In addition, case-control association analysis has shown that the promoter allele of the KLK1 gene containing a poly-G repeat of 12 guanines is associated with hypertensive end stage renal disease (Yu et al, 2002). These findings raise the possibility that the similar poly-G polymorphism at -3432 in the Brn-3c promoter could be functional and hence, a risk factor for late onset hearing loss. To explore this possibility the -3432poly-G polymorphism was functionally characterised and this line of analysis is discussed in the subsequent sections of this chapter.

KLK1 poly-G allele (name)	DNA Sequence	Abbreviation	Similar Brn-3c poly-G allele (abbreviation)	Brn-3c poly-G allele (name)
A	5' ATCGGAGGGGGGGGGGCAATT 3'	$GGA(G)_{10}$	$(G)_{10}$	$(G)_{10}$
B	5' ATCGGAGGCGGGGGGGCAATT 3'	$GGA\ GGC(G)_7$	$GGC(G)_9$	SNPG2
C	5' ATCGGGGGGGGGGGGGCAATT 3'	$(G)_{13}$	$(G)_{13}$	$(G)_{13}$
D	5' ATCGGGGCGGGGGGGGCAATT 3'	$(G)_4C(G)_8$	$(G)_4C(G)_6$	SNPG5
E	5' ATCGGAGGGCGGGGGGCAATT 3'	$GGA\ (G)_3C(G)_6$	$(G)_4C(G)_6$	SNPG5
F	5' ATCGGAGCGGGGGGGGGCAATT 3'	$GGA\ GC(G)_9$	$GGC(G)_9$	SNPG2
G	5' ATCGGAGCGGGGGGGGCAATT 3'	$GGA\ GC(G)_8$	$GGT(G)_8$	SNPG4
H	5' ATCGGAGGGGGGGGGGGCAATT 3'	$GGA\ (G)_{11}$	$(G)_{11}$	$(G)_{11}$
I	5' ATCGGAGGGGGGGGGCAATT 3'	$GGA\ (G)_9$		
J	5' ATCGGAGGGGGGGGCAATT 3'	$GGA\ (G)_8$		
K	5' ATCGGAGGGGGGGGGGGGCAATT 3'	$GGA\ (G)_{12}$	$(G)_{12}$	$(G)_{12}$
M	5' ATCGGAGGGGGGGGGGGGGCAATT 3'	$GGA\ (G)_{13}$	$(G)_{13}$	$(G)_{13}$
P	5' ATCGGAGGGCGGGGGCAATT 3'	$GGA\ (G)_3C(G)_5$	$(G)_4C(G)_6$	SNPG5
Q	5' ATCGGGGGGGGGGGGCAATT 3'	$(G)_{12}$	$(G)_{12}$	$(G)_{12}$
R	5' ATCGGAGGGGGGGGGCAATT 3'	$GGA\ (G)_9CGG$	$(G)_8CG$	SNPG1

Table 5.1 A summary of the poly-G alleles identified at -131 in the KLK1 promoter and comparison with similar poly-G alleles identified at -3432 in the Brn-3c promoter. Poly-G alleles in the KLK1 promoter were identified on a mixed ethnic background; alleles found in Caucasian subjects are highlighted in red (Song et al, 1997; Yu et al, 2002). All poly-G alleles in the Brn-3c promoter were identified in Caucasian subjects.

5.3 The -3432 poly-G polymorphism modulates the binding of OC-2 nuclear proteins.

In order to investigate whether the -3432poly-G polymorphism modifies the binding of nuclear proteins to this region of the Brn-3c promoter EMSA analysis was performed using nuclear extracts from OC-2 cells; an inner ear sensory epithelial cell line which expresses Brn-3c (further information on the OC-2 cell line is discussed in section 5.1.3). Synthetic oligonucleotides spanning the -3432poly-G polymorphism in the Brn-3c promoter were devised to represent the most common alleles and annealed to form double stranded probes (see Table 3.2, method section 3.2.18). For initial analysis two probes were used representing either poly-G allele SNPG1 or SNPG2, the two common alleles that were identified as part of the initial PCR-SSCP screen. To maximise chances of detecting differential protein binding the temperature of the incubation and the amount of nuclear extract added to the binding reaction were varied (see Fig. 5.1).

Lanes 1-3 (Fig. 5.1) show free probe in the absence of nuclear extract; no retarded bands are visualised. In the presence of nuclear extract a complex banding pattern is observed as binding of nuclear proteins to free probe forms protein-DNA complexes with a retarded migration compared to free probe alone (Fig. 5.1, comparing lanes 4-15 with 1-3). At room temperature a multi-banded shifted protein-DNA complex of low mobility, labelled B (Fig. 5.1) and a shifted protein-DNA complex of greater mobility, labelled E (Fig. 5.1) are observed at analogous positions on both SNPG1 and SNPG2 probes but, they are less prominent on probe SNPG2 compared to probe SNPG1 and only clearly visible on probe SNPG1 when a high amount of nuclear extract is used (Fig. 5.1 comparing lanes 7 and 8 with 4). This is in addition to several complexes of lower mobility, which appear common to both probes. It is possible that the same protein(s) is binding to SNPG1 and SNPG2 probes to form complex B and a different protein of lower molecular weight is binding to both probes to form complex E. However, it is important to realise that two different proteins could give identical mobility shifts depending on the nature of their respective protein-DNA complex conformations. In addition, a shifted protein-DNA complex does not necessarily consist of one protein bound to DNA. Proteins can bind DNA sequences as monomers, homodimers or heterodimers; a shifted band observed upon EMSA analysis could represent any of these. All these factors must be taken into consideration when interpreting data generated through EMSA analysis.

In contrast to the observations at room temperature, when the binding reactions are performed on ice, a shifted protein-DNA complex, labelled A (Fig. 5.1) is observed on probe SNPG1 but not probe SNPG2. Instead a complex of greater mobility, labelled C (Fig. 5.1) is observed on the SNPG2 probe but not probe SNPG1 (Fig. 5.1 comparing lanes 10 and 11). Increasing the protein concentration intensifies the formation of complex A on probe SNPG1 and results in the formation of complex C, albeit to less extent (Fig. 5.1 comparing lanes 10 and 13). In contrast, although increasing the protein concentration intensifies the formation of complex C on probe SNPG2, complex A is not observed (Fig. 5.1 comparing lanes 11 and 14). Neither complex A or C are observed on SNPG1 or SNPG2 probes if the binding reaction is carried out at room temperature (shifted protein-DNA complex labelled D is discussed in section 5.3.1.2, page 181 and 5.6, page 210).

Taken together, these results show a complex banding pattern on SNPG1 and SNPG2 probes that is clearly temperature-dependent and subject to the amount of nuclear extract used. Moreover, they suggest that the different poly-G alleles SNPG1 and SNPG2 modulate the binding of nuclear proteins to this region of the Brn-3c promoter.

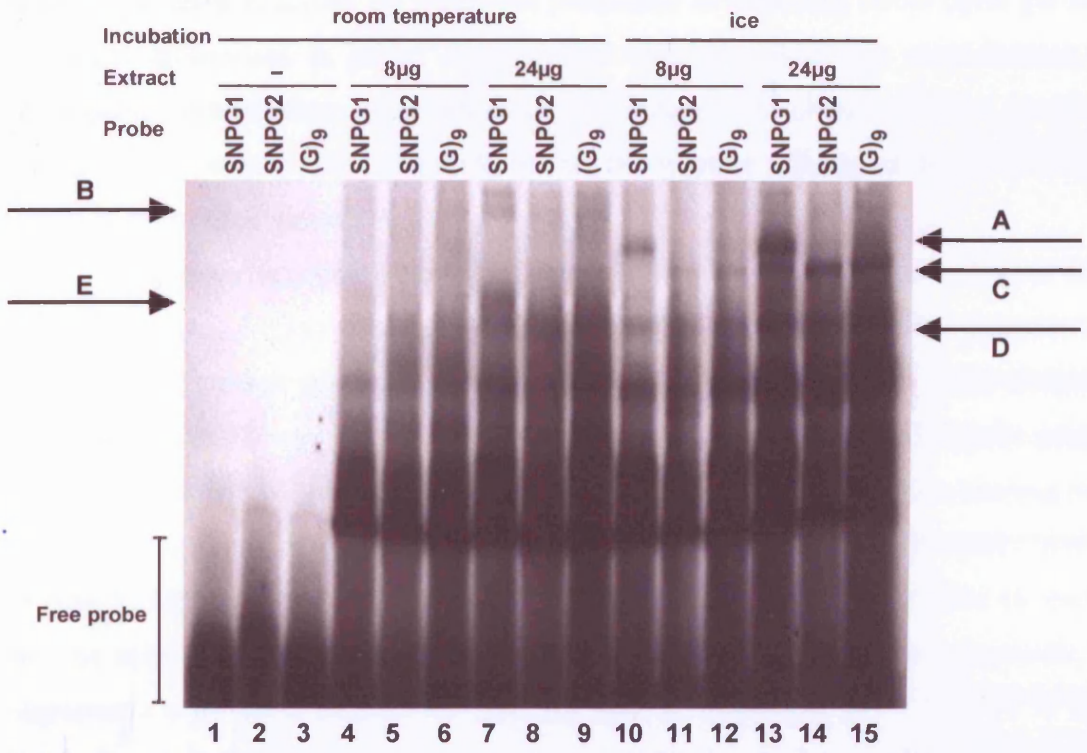


Figure 5.1 Differential protein binding to -3432poly-G alleles SNPG1 and SNPG2. Radiolabelled oligonucleotides representing poly-G alleles SNPG1 and SNPG2 were incubated on ice or at room temperature in the presence of increasing amounts of OC-2 nuclear extract as indicated. In the absence of nuclear extract no retarded bands are visualised (lanes 1-3). In the presence of nuclear extract (lanes 4-15) shifted protein-DNA complexes are observed and are denoted by the arrows: A, B, C, D and E. [$(G)_9$ was later determined to be a sequencing error and not a true allele].

Initial EMSA analysis on the -3432poly-G polymorphism shows differential protein binding to alleles SNPG1 and SNPG2 notably formation of nuclear protein-DNA complex A is only observed on probe SNPG1 and not probe SNPG2 (see Fig. 5.1 lanes 10 and 11, 13 and 14). To investigate this further radiolabelled probes of the additional common poly-G alleles were incubated with OC-2 nuclear extracts (Fig. 5.2a). This approach showed a protein-DNA complex is observed on the additional poly-G probes, including probe SNPG2, at an analogous position to that of complex A on probe SNPG1; presumably all probes are binding the same protein. However, the intensity of complex A appears greater for probe SNPG1 compared to the other poly-G probes (Fig. 5.2a comparing lane 1 across all lanes). This is particularly evident when comparing the intensity of complex A on probes SNPG1 and SNPG2 (Fig. 5.2a comparing lane 1 with 2) suggesting that the nuclear protein or proteins (for simplicity one protein will be assumed to bind hereafter referred to as nuclear protein A) involved in the formation of complex A has a greater affinity for allele SNPG1 compared to the other poly-G alleles examined. However, a shortcoming of EMSA analysis is that there is no internal control for gel loading. When differences in band intensity are observed between two probes, or more such as observed here for nuclear protein A (Fig. 5.2a comparing band A across all lanes) the possibility of pipetting errors upon gel loading or slight differences in probe concentration must be taken into consideration when attempting quantitative interpretations. Therefore, to obtain further clarification competition analysis was performed to test the relative affinity of alleles SNPG1 and SNPG2 for nuclear protein A.

Increasing amounts of unlabelled probe SNPG1 (Fig 5.2b, lanes 2-4) or SNPG2 (Fig. 5.2b, lanes 5-7) to compete with SNPG1 probe for binding nuclear protein A were examined. If nuclear protein A has a greater affinity for the SNPG1 allele compared to SNPG2 one would expect a lower molar excess of unlabelled SNPG1 probe compared to unlabelled SNPG2 probe to compete with labelled SNPG1 probe for binding nuclear protein A. If the affinity of nuclear protein A for allele SNPG1 is particularly high even a high molar excess of unlabelled SNPG2 probe (for example, 1000-fold excess) may not be able to compete with labelled SNPG1 probe for binding nuclear protein A. In agreement with these predictions this approach revealed that allele SNPG1 is able to strongly compete for nuclear protein A; the shifted complex is abolished at 25-fold excess (Fig5.2b, comparing lanes 1 and 2). In contrast allele SNPG2 is unable to induce any significant reduction of the shifted complex at this level (Fig 5.2b, lanes 1 and 5). Even at 100-fold excess, the unlabelled SNPG2 probe showed only partial competition

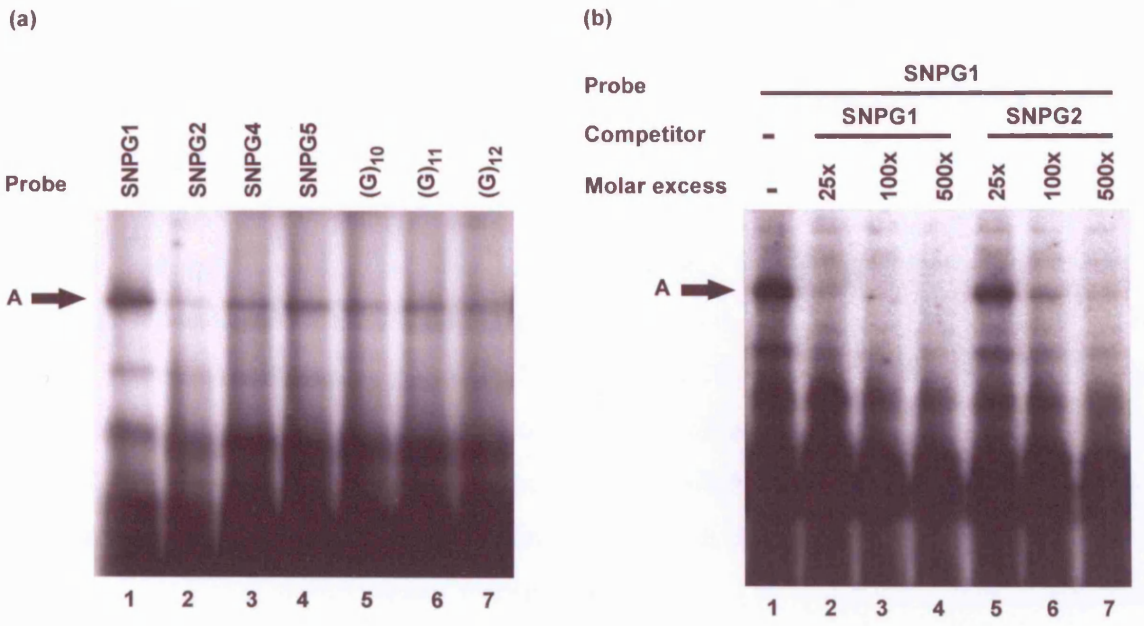


Figure 5.2 The SNPG1 allele of the -3432poly-G polymorphism has a higher affinity for OC-2 derived nuclear protein A than allele SNPG2. The location of nuclear protein-DNA complex A is denoted by the arrow. **(a)** Radiolabeled oligionucleotides of the common poly-G alleles were incubated on ice with 8-10μg of OC-2 nuclear extract. **(b)** The SNPG1 probe was incubated with 8-10μg of OC-2 nuclear extract in the absence of competitor (lane 1), or in the presence of increasing amounts of either unlabeled SNPG1 (lanes 2-4) or unlabeled SNPG2 (lanes 5-7).

that is still not complete at 500-fold excess (Fig 5.2b, comparing lanes 1, 6 and 7). These results confirm that nuclear protein A has a greater affinity for allele SNPG1 compared to SNPG2 and indicate that the interaction of nuclear protein A with allele SNPG1 is sequence specific (the relative affinity of the additional poly-G alleles for nuclear protein A is discussed further in section 5.3.2, page 186).

5.3.1 Identification of nuclear protein A that exhibits differential binding to the -3432 poly-G polymorphic site.

In an attempt to identify nuclear protein A MatInspector software (www.genomatix.de/products/MatInspector/index.html) was used to screen poly-G allele SNPG1 and all common poly-G alleles for putative transcription factor binding sites. MatInspector is a very useful bioinformatic tool to search any specific DNA sequence for potential transcription factor binding sites (Quandt et al, 1995; Cartharius et al, 2005). It draws on a large database of known consensus transcription factor binding sites and nucleotide deviations from these as a set of weight matrices. Sequence similarity between the input sequence and the consensus binding site for a given transcription factor is recorded as a series of two scores: a *core* similarity and a *matrix*

similarity score. The former is "the score of the highest conserved positions of a matrix match" and the latter is "the score of the complete matrix match" (Cartharius et al, 2005). Analysis performed using this software cannot be taken as definitive evidence that a transcription factor binds a particular DNA sequence but rather it provides a guide to the likelihood of transcription factor binding.

The results of this analysis did not reveal any transcription factor binding sites that exhibited a unique matrix similarity to the SNPG1 allele sequence compared to the other poly-G allele sequences (see Appendix C). Therefore, candidate transcription factors chosen for further investigation were based on high matrix similarity of the SNPG1 allele sequence and the additional poly-G allele sequences to known consensus sequences. Two candidates were selected with high matrix similarity scores to the -3432poly-G polymorphic region: stimulating protein 1 (SP1) and early growth response factor-1 (Egr-1/Krox-24/NGFI-A). The transcription factor core promoter binding protein (CPBP/ZBPF/ZF9/KLF6) was also selected, as this exhibited a matrix similarity to the -3432poly-G polymorphic region but not the SNPG1 allele sequence (see Appendix C). Interestingly, SP1 a well-characterised transcription factor known to bind to GC-rich elements (for review see Kaczynski et al, 2003) exhibited a high matrix similarity across all poly-G allele sequences examined with the exception of allele SNPG4 (see Appendix C).

In the first instance, in an attempt to identify nuclear protein A competition analysis was performed using synthetic oligonucleotides for each of the following consensus transcription factor binding sites: SP1, Egr and CPBP (for sequence see Table 3.2 method section 3.2.18). The SNPG1 probe was incubated with OC-2 nuclear protein extract in the presence of increasing amounts of: the unlabelled CPBP consensus sequence (Fig. 5.3, lanes 10 and 11), the unlabelled SP1 consensus sequence (Fig. 5.3, lanes 12 and 13) or the unlabelled Egr consensus sequence (Fig. 5.3, lanes 14 and 15). This approach showed that nuclear protein A is strongly competed from probe SNPG1 by 25-fold excess of the unlabelled SP1 consensus sequence (Fig. 5.3, comparing lanes 1 and 12). In comparison, little reduction in band intensity is observed if the competitor is changed to 25-fold excess of either the unlabelled CPBP or Egr consensus sequence (Fig. 5.3, comparing lanes 1, 10, and 14, respectively). This data suggests that nuclear protein A is a member of the SP1 transcription factor family. Indeed, even at 200-fold excess, the Egr consensus sequence is unable to significantly compete indicating that a member of this transcription factor family is unlikely to be nuclear protein A (Fig. 5.3, lane 15). The consensus sequence for CPBP on the other hand does show some binding

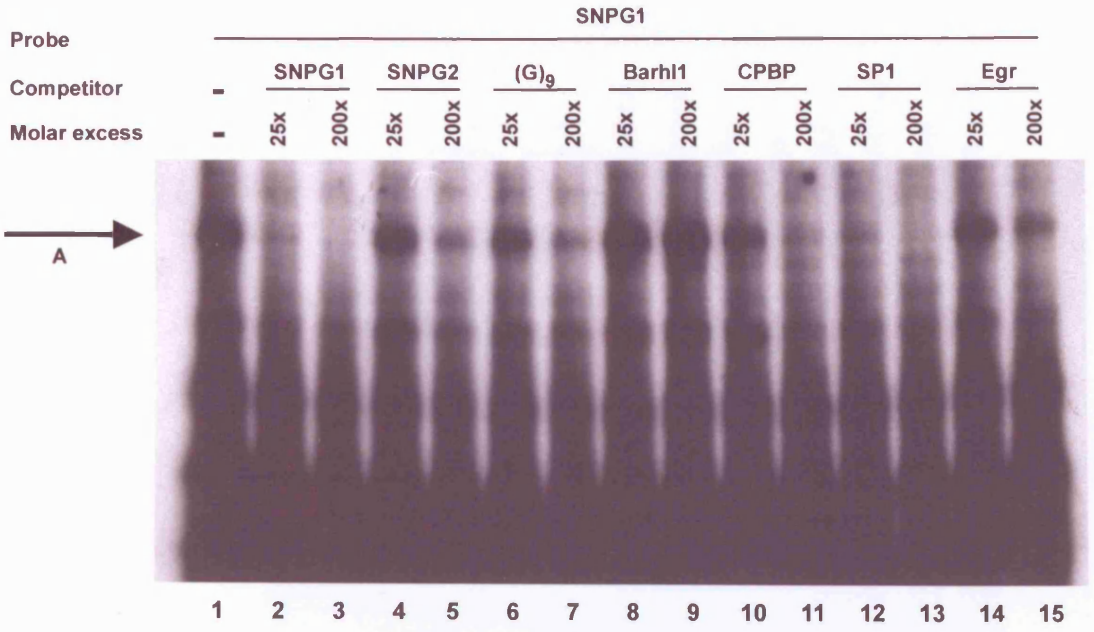


Figure 5.3 Competition analysis suggests that nuclear protein A is a member of the SP1 transcription factor family. The SNPG1 probe was incubated with 8-10μg of OC-2 nuclear extract in the absence of competitor (lane 1) or in the presence of a molar excess of either unlabelled SNPG1 (lanes 2 & 3), SNPG2 (lanes 4 & 5), $(G)_9$ (lanes 6 & 7), unrelated sequence: a section of exon 1 from the Barhl1 gene (lanes 8 & 9), CPBP consensus sequence (lanes 10 & 11), SP1 consensus sequence (lanes 12 & 13) or Egr consensus sequence (lanes 14 & 15). The location of nuclear protein-DNA complex A is denoted by the arrow. [$(G)_9$ was later determined to be a sequencing error and not a true allele].

affinity for nuclear protein A being able to compete at 200-fold excess (Fig. 5.3, lane 11). However, this may be expected since CPBP, a Krüppel-like factor, belongs to a distinct subset of the SP1-like/KLF family and is capable of binding to SP1 sites (Ratziu et al, 1998). The failure of 25- or 200-fold excess of an unrelated sequence from the Barhl1 gene to compete with SNPG1 probe for binding nuclear protein A confirms the sequence specificity of this interaction (Fig. 5.3, lanes 8 and 9).

5.3.1.1 Effect of SP1 consensus sequence as probe in EMSA assay.

To explore the possibility that nuclear protein A is a member of the SP1 transcription factor family further EMSA analysis was performed using the SP1 consensus sequence as the probe. In the absence of any competitors this approach showed binding of OC-2 nuclear proteins to the SP1 consensus sequence produces a near-identical pattern of nuclear protein-DNA complex formation to that obtained with the SNPG1 probe (Fig. 5.4, comparing lanes 2 and 3). In particular, the location and intensity of nuclear protein-DNA complex A suggests that both the SP1 consensus and SNPG1 probes are binding the same protein (Fig. 5.4, comparing lanes 2 and 3). Competition analysis is in

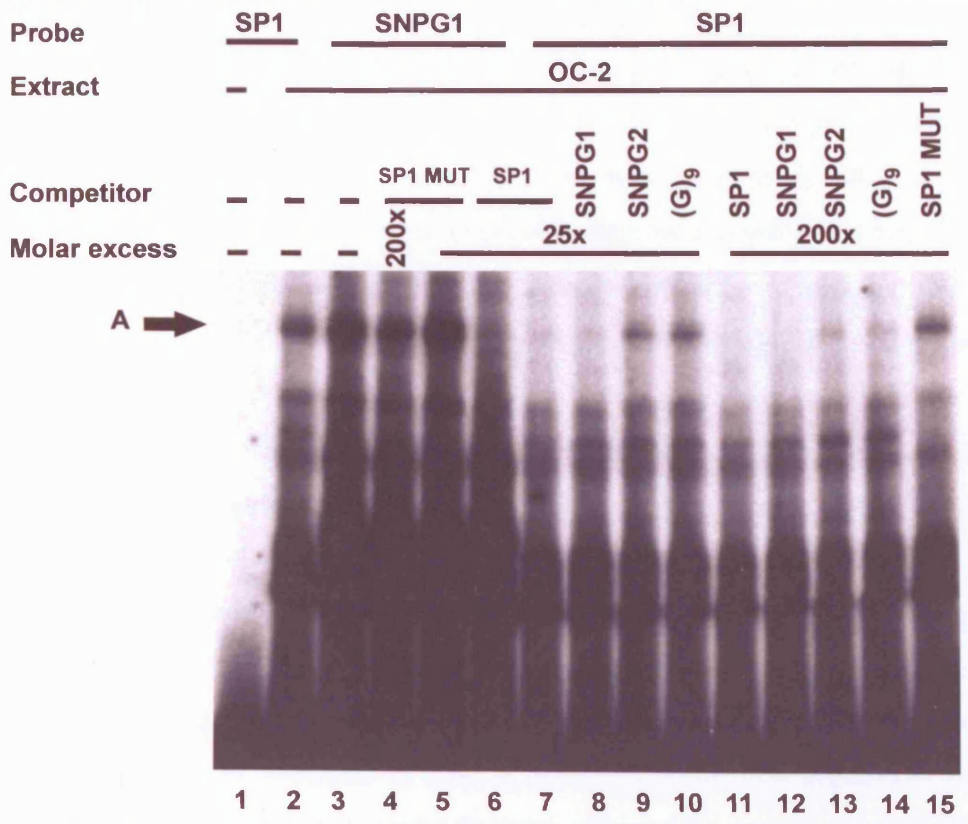


Figure 5.4 The affinity of poly-G allele SNPG1 for nuclear protein A is comparable to that of the SP1 consensus sequence. The labelled SP1 consensus sequence (lanes 1-2 & 7-15) or the SNPG1 probe (lanes 3-6) were incubated in the absence of nuclear extract (lane 1) or with 8-10μg of OC-2 nuclear extract (lanes 2-15) in the absence of competitor (lanes 1-3) or in the presence of a molar excess of either unlabelled oligonucleotide containing a mutated SP1binding site (lanes 4-5 & 15), SP1 consensus sequence (lanes 6-7 & 11), SNPG1 (lanes 8 & 12), SNPG2 (lanes 9 & 13), or $(G)_9$ (lanes 10 & 14). The location of nuclear protein-DNA complex A is denoted by the arrow. [$(G)_9$ was later determined to be a sequencing error and not a true allele].

agreement with this suggestion. The SP1 consensus sequence and allele SNPG1 are both able to significantly compete nuclear protein A from the consensus SP1 binding site at 25-fold competition (Fig. 5.4, comparing lanes 2, 7 and 8). In contrast, allele SNPG2 exhibits much weaker competition with shifted complex A still apparent even at 200-fold competition (Fig. 5.4, comparing lanes 2, 9 and 13). Failure of an unlabelled oligonucleotide containing a mutated SP1 binding site to compete nuclear protein A from either the SP1 consensus sequence (Fig. 5.4, comparing lanes 2 and 15), or the SNPG1 probe (Fig. 5.4, comparing lanes 3, 4 and 5) confirms the specificity of this interaction, in that it is dependent on the presence of a functional SP1 binding site.

Taken together these results confirm that both the SP1 consensus sequence and allele SNPG1 bind nuclear protein A with comparable affinities. Moreover, they suggest allele SNPG1 constitutes a high affinity SP1 binding site and that the affinity of this site for SP1 is modulated by the sequence variation present in the other -3432poly-G alleles.

5.3.1.2 Effect of Zinc on binding of nuclear protein A.

SP1 is a zinc finger transcription factor containing a highly conserved DNA-binding domain at the C-terminus characterised by three tandem $(cystein)_2(histidine)_2$ zinc finger motifs (for reviews see Cook et al, 1999; Kaczynski et al, 2003). Cys_2His_2 zinc finger proteins such as SP1 use zinc to provide structural stability for sequence specific DNA binding; each zinc finger is tethered to a zinc ion via two Cys and two His residues such that the finger folds into a compact $\beta\beta\alpha$ domain (for review see Wolfe et al, 2000).

To assess whether nuclear protein A is responsive to the presence of zinc ions in the EMSA binding reaction as would be expected for a member of the SP1 transcription factor family a zinc titration was performed in the EMSA binding assay (see Fig. 5.5). This approach showed that binding of nuclear protein A is strongly augmented at comparable amounts on SNPG1 and SP1 consensus probes as the concentration of zinc in the binding reaction is increased (Fig. 5.5 comparing lanes 1 and 2, 5 and 6, and 9 and 10). Additionally, presence of zinc in the binding reaction appears to promote the binding of nuclear protein A to probe SNPG2, although to a much less extent compared to either SNPG1 or SP1 consensus probes (Fig. 5.5 comparing lanes 3, 7 and 11). These results are consistent with previous observations that allele SNPG2 has a reduced affinity for nuclear protein A compared to allele SNPG1, and that allele SNPG1 and the SP1 consensus sequence exhibit comparable affinities for nuclear protein A. Moreover, they suggest that nuclear protein A is a zinc finger transcription factor; requiring zinc as a cofactor for enhanced binding.

Binding of SP1 to its recognition sequences can be abolished by zinc depletion (Westin et al, 1988). With this in mind it may have been interesting to include a chelating agent such EDTA in the binding assay to assess whether zinc ion depletion is consistent with absence of nuclear protein A binding. Although, the standard Parker buffer used here in these EMSA assays contains EDTA, this is at a low concentration (1mM) compared to those used for metal depletion (Westin and Schaffner, 1988). However, the results of the zinc titration, as shown in Fig. 5.5 are consistent with the notion that nuclear protein A is a member of the SP1 transcription factor family. The results of this zinc titration also revealed nuclear protein-DNA complex C and an additional nuclear protein-DNA complex, termed D, to be affected by the presence of zinc in the EMSA binding reaction (see arrows C and D, respectively in Fig. 5.5). These results are discussed further in section 5.3.3, page 193 and section 5.6, page 210, respectively.

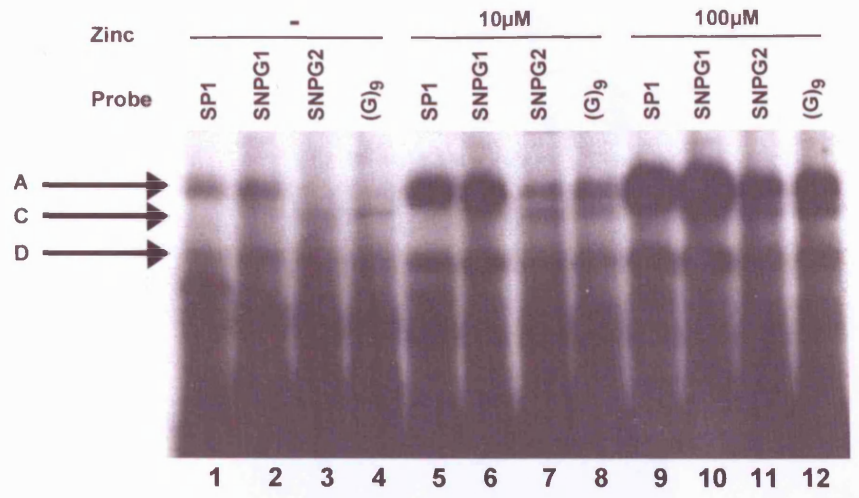


Figure 5.5 Effect of Zinc on binding of nuclear protein A. The labelled SP1 consensus sequence (lanes 1, 5 & 9), SNPG1 probe (lanes 2, 6 & 10), SNPG2 probe (lanes 3, 7 & 11) or $(G)_9$ probe (lanes 4, 8 & 12) were incubated with 8-10μg of OC-2 nuclear extract in the absence of zinc (lanes 1-4) or in the presence of 10μM zinc (lanes 5-8) or 100μM zinc (lanes 9-12). Arrows A, C & D denote shifted protein-DNA complexes that are affected by the presence of zinc in the EMSA binding reaction. [$(G)_9$ was later determined to be a sequencing error and not a true allele].

5.3.1.3 Supershift analysis to identify nuclear protein A.

In an attempt to confirm that OC-2 derived nuclear protein A is SP1, supershift analysis was performed by incubating a mouse monoclonal anti-SP1 antibody [SP1 (1C6) X; Santa Cruz Biotechnology, Inc.] with OC-2 nuclear extracts subsequent to addition of either the SNPG1 or SP1 consensus probe (Fig. 5.6a). If nuclear protein A is SP1 and the epitope to which the anti-SP1 antibody was raised is exposed upon binding of nuclear protein A to DNA probe one would expect the anti-SP1 antibody to bind forming one large multimeric complex. Multimeric complexes such as these possess retarded mobility compared to protein-DNA complex alone and can be visualised by the presence of a novel band high on the gel, termed a supershift. However, this approach did not induce a supershift of nuclear protein-DNA complex A on either the SNPG1 or SP1 consensus probes (Fig. 5.6a comparing lanes 1 and 2 with 3 and 4). Consequently, the supershift analysis was varied by incubating the anti-SP1 antibody prior to addition of labelled probe. If the anti-SP1 antibody binds to a site on nuclear protein A that is essential for DNA binding this could block the ability of nuclear protein A to bind to labelled probe and result in the disappearance of nuclear protein-DNA complex A. Additionally, several attempts were made to optimise the supershift analysis by varying the temperature and length of the incubation when the anti-SP1 antibody was added prior to the probe (4°C O/N, on ice for 30 minutes or on ice for 2hrs) and following addition of labelled probe (on ice for 30 minutes). However a positive result was not

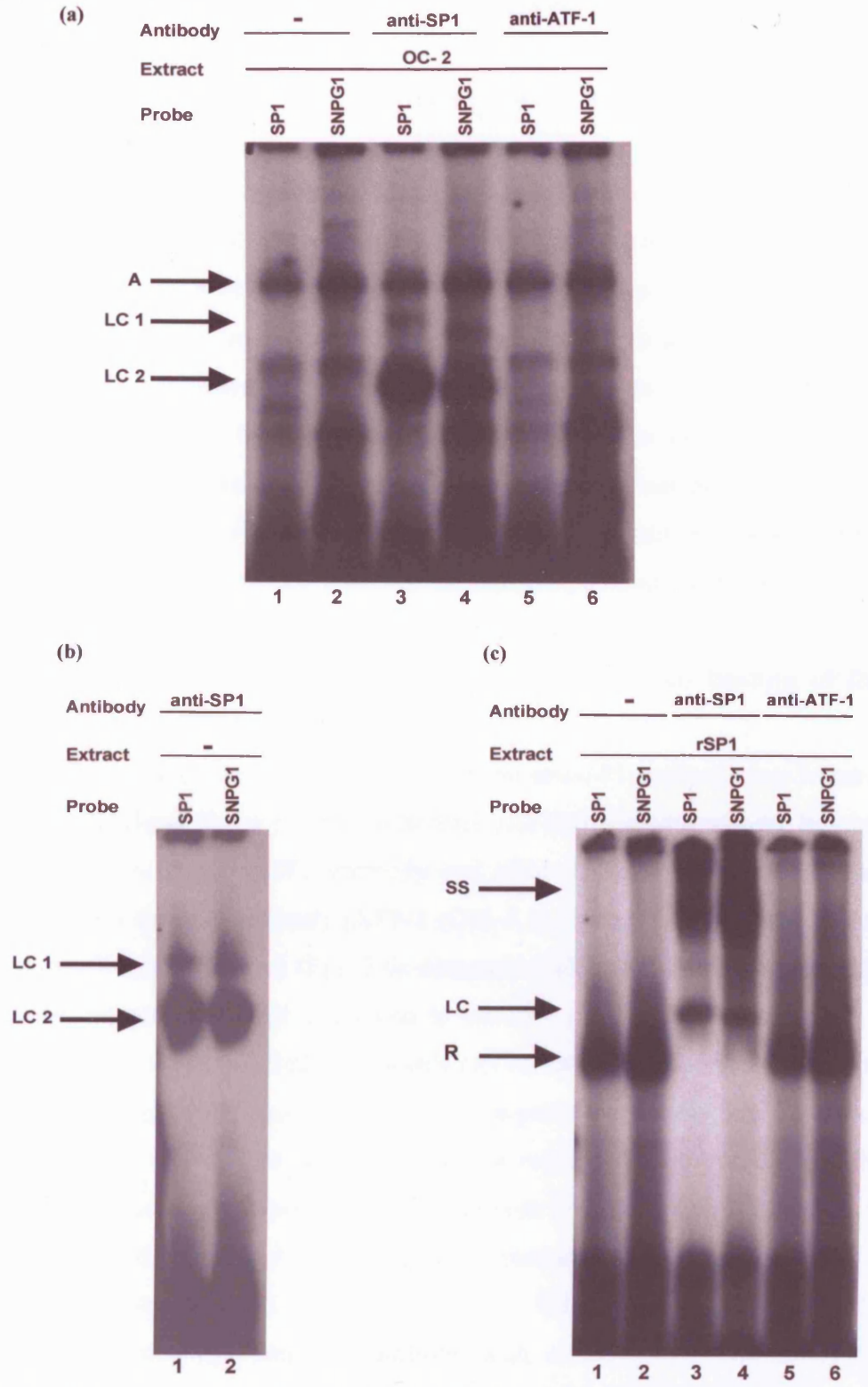


Figure 5.6 Supershift analysis to identify nuclear protein A. (a) The labelled SP1 consensus sequence or SNPG1 probe were incubated with 8-10μg of OC-2 nuclear extract in the absence of antibody (lanes 1-2) or in the presence of anti-SP1 antibody (lanes 3-4) or anti-ATF-1 antibody (lanes 5-6) as indicated. The location of nuclear protein-DNA complex A is denoted by the arrow, A. **(b)** The labelled SP1 consensus sequence (lane 1) or the SNPG1 probe (lane 2) were incubated with anti-SP1 antibody in the absence of OC-2 nuclear extracts. **(a)** and **(b)** In the presence of the anti-SP1 antibody lower complexes LC1 & LC2 can be observed lower down the gel. **(c)** The labelled SP1 consensus sequence or SNPG1 probe were incubated with 258ng recombinant SP1 protein in the absence of antibody (lanes 1-2) or in the presence of anti-SP1 antibody (lanes 3-4) or anti-ATF-1 antibody (lanes 5-6) as indicated. R denotes the complex formed by the binding of recombinant SP1 protein to labelled probe. SS denotes supershift of recombinant SP1 protein. In the presence of the anti-SP1 antibody lower complexes (LC) are observed. In each case **(a), (b)** and **(c)** antibody was added subsequent to addition of probe and reactions were allowed to incubate for a further 30 minutes on ice.

obtained under any of these conditions. It is possible that these results are due to the fact that nuclear protein A is not SP1. It could be a different member of the SP1 transcription factor family; 17 SP1-like/KLF genes have been identified in mouse (the OC-2 cell line used for preparation of nuclear extract is derived from Immortomouse; see section 5.1.3) (for review see Kaczynski et al, 2003). Alternatively, nuclear protein A may belong to a different transcription factor family or be a novel transcription factor, albeit one that exhibits high affinity for the SP1 consensus sequence. On the other hand, it should also be considered that these results are due to unsuccessful supershifts and hence, nuclear protein A is SP1. Certainly, the epitope to which the anti-SP1 antibody was raised may not be not exposed on nuclear protein A and thus, accessible for binding. It may be masked by the presence of other nuclear proteins within the extract or the conformation of nuclear protein A under the conditions of these EMSA assays may not expose the epitope site in a favourable position for binding of the anti-SP1 antibody. Alternatively, it is possible that the nuclear extract buffer contains a component or components that have an adverse effect on binding of the anti-SP1 antibody to nuclear protein A.

It is notable that in the presence of the anti-SP1 antibody two lower complexes (termed LC 1 and 2) are present on SNPG1 and SP1 consensus probe that are absent in the absence of the anti-SP1 antibody and when a mouse monoclonal anti-activating transcription factor-1 antibody [ATF-1 (C41-5.1) X, Santa Cruz Biotechnology, Inc.] is used as a negative control (Fig. 5.6a comparing all lanes; LC1 having a slightly lower mobility on probe SNPG1 compared to the SP1 consensus probe). Lower complexes such as these that are shifted downwards can be induced when a specific antibody binds a nuclear protein and removes it from a multi-protein complex bound to labelled probe. However, the intensity of complex A is not reduced in the presence of the anti-SP1 antibody as one might expect if a nuclear protein was being removed from a complex of proteins bound to labelled probe. Indeed, further analysis suggests that LC 1 and 2 are the result of the anti-SP1 antibody binding to the SNPG1 and SP1 consensus probes direct; incubating the anti-SP1 antibody with either the SNPG1 or SP1 consensus probes in the absence of OC-2 nuclear extracts results in formation of complexes LC 1 and 2 on both probes (Fig. 5.6b lanes 1 and 2). Hence, the supershift analysis appears to be confounded by the fact that free probe has the ability to bind the anti-SP1 antibody. To counteract the possible effect of free probe sequestering the anti-SP1 antibody the supershift analysis was repeated with higher concentrations of OC-2 nuclear extracts (24μg), but this approach was not successful.

In the absence of a supershift with the anti-SP1 antibody, EMSA experiments were performed with affinity purified recombinant human SP1 protein (Promega) in place of OC-2 nuclear extracts in order to establish whether the -3432poly-G polymorphism does in fact modulate a high affinity SP1 binding site. The results of this analysis showed that recombinant SP1 protein can bind SNPG1 and SP1 consensus probes and this is discussed in greater detail in the next section, 5.3.2. Having shown that recombinant SP1 protein can bind SNPG1 and SP1 consensus probes the supershift analysis was repeated using recombinant SP1 protein in place of the OC-2 nuclear extracts to assess whether the anti-SP1 antibody used in supershift analysis was functional; use of a non-functional antibody could account for the lack of a positive supershift of nuclear protein A (Fig. 5.6c). However, from this analysis it is clear that the anti-SP1 antibody used can induce a complete supershift of recombinant SP1 protein whether bound to the SP1 consensus or SNPG1 probes (Fig. 5.6c lanes 3 and 4) ruling out the possibility that the anti-SP1 antibody used in these assays is non-functional. However, the failure of the SP1 supershift analysis to produce positive results with the OC-2 nuclear extracts despite the use of an anti-SP1 antibody that was confirmed to be functional does not mean that nuclear protein A is not SP1. There are many reasons why the SP1 supershift analysis could have failed with the OC-2 nuclear extracts as has been discussed. Importantly, western immunoblot analysis shows that the OC-2 cell line expresses SP1 protein (see Fig. 5.12, section 5.4, page 201). Hence, in light of this observation one would expect the SP1 consensus probe to bind endogenous SP1 in these EMSA assays. Consequently, the absence of a supershift on the SP1 consensus probe suggests that the results obtained probably reflect a failure of the antibody to bind endogenous SP1 under the assay conditions used.

5.3.2 The -3432 poly-G polymorphism creates a high affinity SP1 binding site.

In the absence of a supershift with the SP1 antibody and in order to establish whether the -3432poly-G polymorphism does in fact modulate a high affinity SP1 binding site EMSA analysis was performed using affinity purified recombinant human SP1 protein (Promega) in place of OC-2 nuclear extracts. Use of purified protein in EMSA analysis can lead to spurious results through non-specific binding of the highly abundant purified protein. However, this approach showed that recombinant SP1 protein is able to bind probe SNPG1 to form a shifted protein-DNA complex that appears equivalent in

intensity, if not greater, to that formed on the SP1 consensus probe and importantly, much greater than that observed on probe SNPG2 (Fig. 5.7a, complex R in lanes 4, 8 and 12). These observed differences in binding intensity to SNPG1 and SNPG2 probes presumably reflect differences in binding affinity between these two poly-G alleles for recombinant SP1 protein; they are similar to those observed with OC-2 nuclear extracts (see Fig. 5.7a, lanes 6 and 10, respectively for OC-2 nuclear extracts). It is notable that the complex formed by the binding of recombinant SP1 protein has greater mobility compared to that formed by the binding of OC-2 derived nuclear protein A to the same probes (Fig. 5.7a, comparing lanes 4, 8 and 12 with 2, 6 and 9). However, recombinant SP1 produced in the baculovirus expression system does not undergo the same post-translational modifications as endogenous SP1 (molecular weight of 57kDa / 82kDa for recombinant SP1 to 95 kDa / 106kDa for mammalian SP1, Promega, U.K) and is not expected to have the same mobility.

Competition experiments show that the unlabelled SNPG1 probe is able to compete recombinant SP1 protein from its consensus binding site with an affinity similar to that of the unlabeled SP1 consensus sequence itself; strong competition is evident at 25-fold excess that progresses to the disappearance of the shifted complex at 200-fold competition (Fig. 5.7b, comparing lanes 1, 2, 3 with 9 and 10). This suggests that allele SNPG1 is a high affinity SP1 binding site not dissimilar to the consensus SP1 binding site. Furthermore, it is clear from this analysis that allele SNPG2 is a much lower affinity SP1 binding site; only partial competition is evident at 200-fold excess (Fig. 5.7b, lane 11). This is a similar level of competition exerted by the consensus sequences for Egr and CPBP (Fig. 5.7b, comparing lanes 1, 11, 14 and 15) both of which are known to have a low affinity for SP1 (Ebert and Wong, 1995; Ratziu et al, 1998). Failure of an unlabelled oligonucleotide containing a mutated SP1 binding site to compete recombinant SP1 protein from its consensus binding site, even at 200-fold excess (Fig. 5.7b, lane 13) confirms the sequence specificity of the interaction between recombinant SP1 protein and the consensus SP1 binding site in that it is dependant on the presence of a functional SP1 binding site and further suggests that the interaction between poly-G alleles and recombinant SP1 protein is likewise dependent on the presence of a functional SP1 binding site.

EMSA analysis using either recombinant SP1 protein or OC-2 nuclear extracts was extended to the rest of the poly-G alleles in order to investigate their relative affinity for SP1. The ability of each poly-G allele to compete with the consensus SP1 binding site for either recombinant SP1 protein (Fig. 5.8a) or nuclear protein A

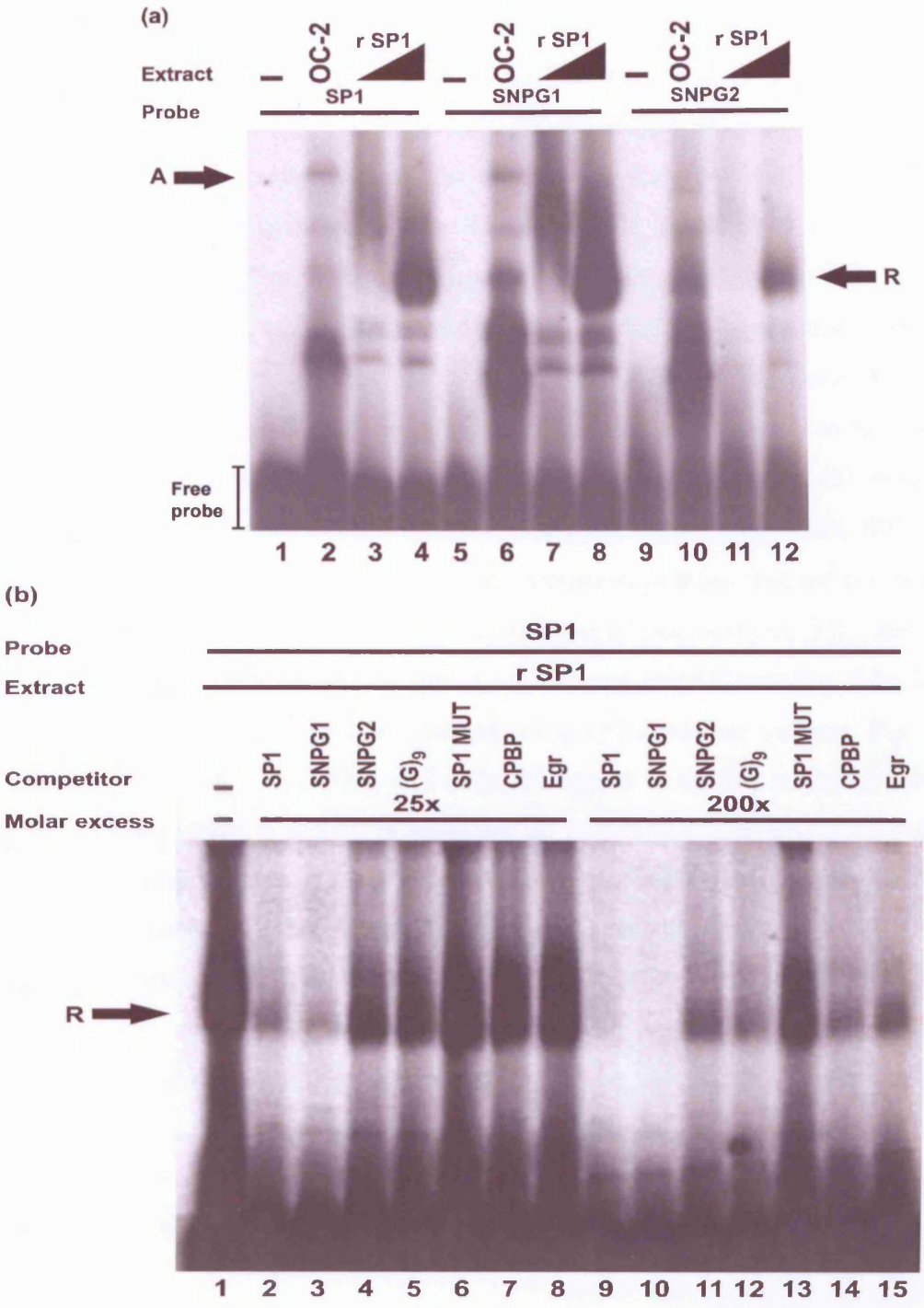


Figure 5.7 The -3432 poly-G polymorphism generates a high affinity binding site for SP1. (a) The labelled consensus SP1 sequence (lanes 1-4), SNPG1 probe (lanes 5-8) or SNPG2 probe (lanes 9-12) were incubated in the absence of extract (lanes 1, 5 & 9), in the presence of 8-10μg OC-2 nuclear extract (lanes 2, 6 & 10) or in the presence of increasing amounts of recombinant SP1 protein, 129ng (lanes 3, 7 & 11) and 258ng (lanes 4, 8 & 12). Arrow A denotes the location of shifted nuclear protein-DNA complex A. Arrow R denotes the location of the shifted protein-DNA complex formed by the binding of recombinant SP1 to labelled probes. **(b)** The labelled SP1 consensus sequence was incubated with 258ng recombinant SP1 protein in the absence of competitor (lane 1) or in the presence of a molar excess of unlabelled SP1 consensus sequence (lanes 2 & 9), SNPG1 (lanes 3 & 10), SNPG2 (lanes 4 & 11), $(G)_9$ (lanes 5 & 12), oligonucleotide containing a mutated SP1 binding site (lanes 6 & 13), CPBP consensus sequence (lanes 7 & 14) or Egr consensus sequence (lanes 8 & 15). [$(G)_9$ was later determined to be a sequencing error and not a true allele].

(Fig. 5.8b) was examined. Data obtained using recombinant SP1 protein confirmed that allele SNPG1 has the highest affinity SP1 binding site closely followed by allele $(G)_{10}$; both are able to strongly compete at 25-fold excess (Fig. 5.8a, lanes 1, 2 and 6). Whereas, allele SNPG2 and SNPG4 exhibit the lowest affinity for SP1 binding; neither of these unlabeled oligonucleotides appear able to significantly compete at 25-fold excess and only partial reduction of the shifted complex is observed at 100-fold competition (Fig. 5.8a, comparing lanes 1, 3, 4, 10 and 11). This is in agreement with analysis performed using MatInspector software which showed that allele SNPG4 did not exhibit a strong matrix similarity to the SP1 consensus sequence (see section 5.3.1, page 177 and Appendix C). It is possible that this may be due to the fact that allele SNPG4 contains a longer (TGG) motif and consequently is more GT-rich compared to the other poly-G alleles (see Table 4.3, section 4.4.2, page 156); SP1 is known to exhibit binding affinity for GC-rich sites compared to those that are GT-rich (for review see Kaczynski et al, 2003). The three remaining alleles, SNPG5, $(G)_{11}$ and $(G)_{12}$ exhibit intermediate affinity for SP1 to that of SNPG1 and SNPG2 (see Fig. 5.8a, lanes 5, 7 and 8). This data is similar to that obtained using OC-2 nuclear extracts, Fig. 5.8b. Alleles SNPG4, SNPG5, $(G)_{10}$, $(G)_{11}$ and $(G)_{12}$ all appear to exhibit a similar level of affinity for nuclear protein A at a level intermediate to that of alleles SNPG1 and SNPG2. This is particularly evident at 25 fold competition; the additional poly-G alleles moderately compete nuclear protein A from the SP1 consensus binding site at levels intermediate between alleles SNPG1 and SNPG2 which exhibit strong and weak competition, respectively (Fig 5.8b comparing lane 1 across lanes 2-8). Taken together, results obtained using recombinant SP1 protein are consistent with those obtained using OC-2 nuclear extracts and demonstrate that the different poly-G alleles are able to bind recombinant SP1 protein in a highly specific manner that is similar to their respective abilities to bind nuclear protein A. This lends further support to the fact that nuclear protein A is SP1.

In summary, the results from this EMSA analysis show that the -3432poly-G polymorphism in the Brn-3c promoter modifies binding affinity for a specific nuclear protein (protein A) expressed in OC-2 cells. Of the seven common poly-G alleles examined, allele SNPG1 has at least a 4-fold higher affinity compared to some of the other poly- G alleles for nuclear protein A (comparing relative levels of competition in Fig. 5.2b and 5.8b). Specific competition with a SP1 consensus sequence but not an oligonucleotide containing a mutated SP1 binding site or an unrelated sequence suggests that nuclear protein A is a member of the SP1 transcription factor family (Fig.

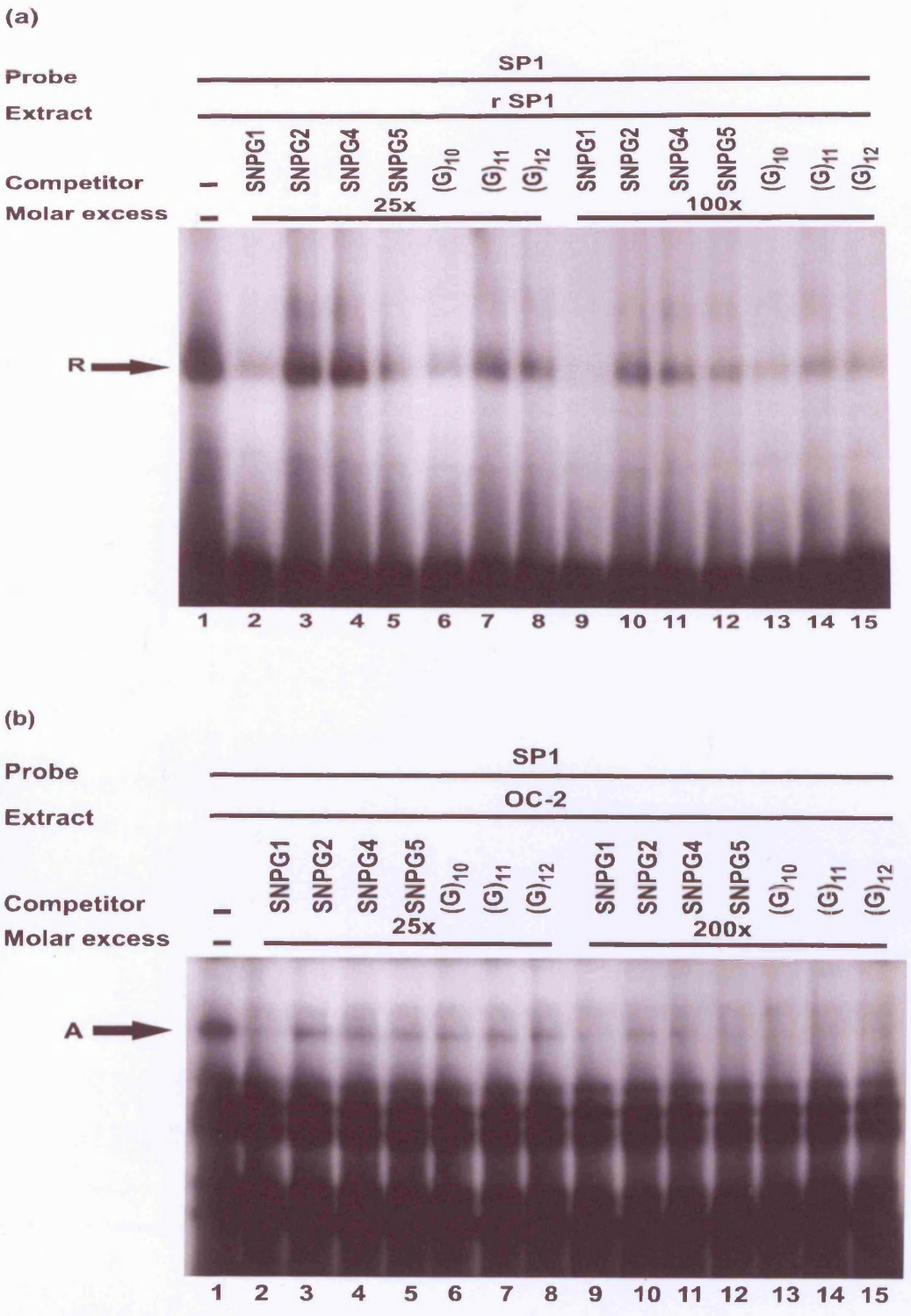


Figure 5.8 The different poly-G alleles bind recombinant SP1 in a highly specific manner that is similar to their respective abilities to bind nuclear protein A. The labelled SP1 consensus sequence was incubated with either **(a)** 258ng recombinant SP1 protein or **(b)** 8-10μg of OC-2 nuclear extract in the absence of competitor (lane 1) or in the presence of a molar excess of unlabelled SNPG1 (lanes 2 & 9), SNPG2 (lanes 3 & 10), SNPG4 (lanes 4 & 11), SNPG5 (lanes 5 & 12), $(G)_{10}$ (lanes 6 & 13), $(G)_{11}$ (lanes 7 & 14) or $(G)_{12}$ (lanes 8 & 15). Arrow A denotes the location of shifted nuclear protein-DNA complex A. Arrow R denotes the location of the shifted protein-DNA complex formed by the binding of recombinant SP1 to labelled probe.

5.3 and 5.4). The finding that the SP1 consensus probe produces a near-identical pattern of protein-DNA complex formation to that observed on probe SNPG1, in particular, the location and intensity of protein-DNA complex A is in agreement with this suggestion (Fig. 5.4). In addition, the enhanced binding of nuclear protein A to SNPG1 and SP1 consensus probes in the presence of zinc is consistent with nuclear protein A being a zinc finger transcription factor such as SP1. Further confirmation that nuclear protein A is SP1 was not obtained from supershift analysis with an anti-SP1 antibody (Fig. 5.6a). However, this does not rule out the possibility that nuclear protein A is SP1. Successful interaction of an antibody with a protein bound to labelled DNA probe (or even antibody-protein interaction prior to addition to labelled DNA probe) is subject to the assay conditions and there are many reasons why the SP1 supershift could have failed as previously discussed. Had time permitted Shift-Western blotting of nuclear protein-DNA complexes formed with the SNPG1 and SP1 consensus probes may have been useful to gain further evidence that nuclear protein A is SP1. However, Shift-Western blotting which combines the EMSA gel shift technique with that of the Western transfer of proteins (immunoblotting) can be notoriously difficult to perform (see Demczuk et al, 1993). Similarly, matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF) a powerful technique that identifies isolated proteins by a combination of accurate mass determination and peptide mass fingerprinting (Chaurand et al, 1999) may have confirmed that nuclear protein A is SP1. MALDI-TOF has the added advantage of being able to discriminate from several proteins in a mixture when the complete isolation of a single protein is not possible (Jenson et al, 1996). However, it was not possible to pursue this line of analysis within the time constraints of the project. In any case, analysis with recombinant SP1 protein shows that recombinant SP1 can bind to the SNPG1 allele sequence to produce a bandshift pattern that is identical to that observed with the SP1 consensus sequence (Fig. 5.7a). Moreover, recombinant SP1 protein appears to bind to allele SNPG1 with an affinity comparable to that of the consensus SP1 binding site (Fig. 5.7b). Furthermore, the relative affinity of recombinant SP1 protein for the SP1 consensus sequence, allele SNPG1 and additional poly-G alleles is very similar to that of nuclear protein A for these sequences (comparing Fig.5.8a and b). Hence, the results obtained using recombinant SP1 protein are in good agreement with those obtained using OC-2 nuclear extracts. Therefore, collectively the data suggest that allele SNPG1 is indeed a high affinity SP1 binding site and are highly indicative that OC-2 derived nuclear protein A that exhibits high binding affinity to allele SNPG1 of the -3432poly-G polymorphism in the Brn-3c promoter is SP1.

5.3.3 Identification of nuclear protein C that exhibits differential binding to the -3432poly-G polymorphic site.

Initial EMSA analysis with OC-2 nuclear extracts revealed a complex banding pattern on -3432poly-G probes SNPG1 and SNPG2 (see Fig. 5.1, page 175). In particular, the complex formed by the binding of nuclear protein A to poly-G probes is more prominent on probe SNPG1 compared to probe SNPG2 whereas, in contrast, a shifted protein-DNA complex termed C is more prominent on probe SNPG2 compared to probe SNPG1 (Fig. 5.1, page 175, comparing lanes 10 and 11 and 13 and 14). Subsequent analysis has confirmed that nuclear protein A has a greater affinity for poly-G allele SNPG1 compared to the other poly-G alleles and suggests that nuclear protein A is the transcription factor SP1 (section 5.3.1 and 5.3.2). The preliminary data presented in Fig 5.1 (page 175) suggests that the nuclear protein or proteins involved in the formation of shifted protein-DNA complex C (hereafter referred to as nuclear protein C) has a greater affinity for poly-G allele SNPG2 compared to allele SNPG1; this is in contrast to the affinity of nuclear protein A for these two poly-G alleles. Hence, raising the possibility that nuclear proteins A and C are distinct.

To explore this possibility analysis was performed using MatInspector software (www.genomatix.de/products/MatInspector /index.html) for putative transcription factor binding sites that exhibited a unique or high matrix similarity to the SNPG2 allele sequence compared to allele SNPG1 (see Appendix C). The results of this analysis showed that the transcription factor Egr-4 (NGF1-C) exhibits a high matrix similarity to the SNPG2 allele sequence compared to allele SNPG1 (see Appendix C). Egr-4 belongs to the Egr transcription factor family, which consists of a group of immediate early response genes that are transiently activated in response to environmental signals (for reviews see O'Donovan et al, 1999; Thiel and Cibelli et al, 2002). In light of these findings and to seek further clarification a limited set of preliminary competition experiments were performed in the EMSA assay as far as was possible to investigate within the time constraints of this project (see Fig. 5.9).

To assess whether the binding of nuclear protein C to poly-G allele SNPG2 is sequence specific and in order to assess whether nuclear protein C may be a member of the Egr transcription factor family the Egr consensus sequence and an oligonucloetide containing a mutated Egr binding site were used as competitors in EMSA analysis (Fig. 5.9a). This line of analysis showed that nuclear protein C is strongly competed from probe SNPG2 by 100- and 500-fold excess of the unlabelled SNPG2 probe (Fig. 5.9a, comparing lanes 1, 7 and 8) and by 500-fold excess of the unlabelled Egr consensus

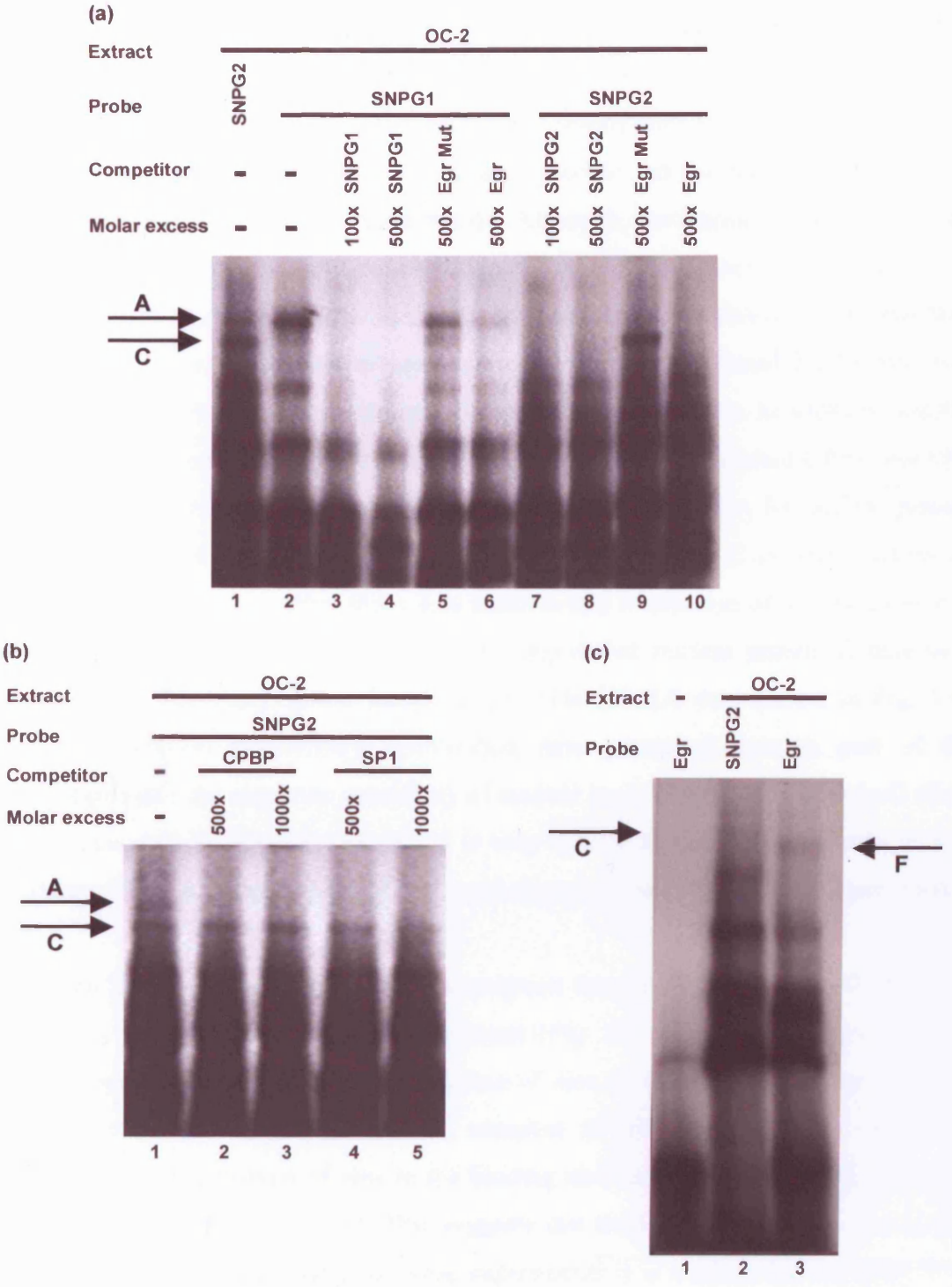


Figure 5.9 EMSA analysis to identify nuclear protein C. (a) The SNPG2 probe (lanes 1, 7-10) or SNPG1 probe (lanes 2-6) were incubated in the presence of 16μg OC-2 nuclear extract in the absence of competitor (lanes 1 & 2) or in the presence of a molar excess of unlabelled SNPG1 (lanes 3 & 4), an oligonucleotide containing a mutated Egr binding site (lanes 5 & 9), Egr consensus sequence (lanes 6 & 10) or SNPG2 (lanes 7 & 8). **(b)** The SNPG2 probe was incubated with 8μg OC-2 nuclear extract in the absence of competitor (lane 1) or in the presence of a molar excess of unlabelled CPBP consensus sequence (lanes 2 & 3) or SP1 consensus sequence (lanes 4 & 5). **(c)** The labelled Egr consensus sequence was incubated either in the absence (lane 1) or in the presence (lane 3) of OC-2 nuclear extract (8μg) and the pattern of shifted nuclear protein-DNA complex formation was compared to labelled SNPG2 probe in the presence of OC-2 nuclear extract, 8μg (lane 2). Arrows A, C and F denote the location of shifted nuclear protein-DNA complexes A, C and F, respectively.

sequence (Fig. 5.9a, lane 10); the shifted complex for nuclear protein C is abolished in each case. In contrast, competition with 500-fold excess of an unlabelled oligonucleotide containing a mutated Egr binding site failed to induce any significant reduction of complex C (Fig. 5.9a, lane 14). These results suggest that the interaction of nuclear protein C with allele SNPG2 is sequence specific and that nuclear protein C is a member of the Egr transcription factor family. Although, the amount of unlabelled Egr consensus sequence used, as competitor DNA in this EMSA experiment is high (500-fold excess) and could arguably lead to non-specific binding. However, it is clear that 500-fold excess of an unlabelled oligonucleotide containing a mutated Egr binding site cannot compete at this level. Hence, suggesting this is not the case. In addition, similar results were obtained when 500- and 1000-fold excess of the unlabelled CPBP and SP1 consensus sequences were used as competitors; the shifted complex for nuclear protein C is not abolished in either case (Fig. 5.9b comparing complex C in lane 1 across all lanes). Therefore, taken together these data confirm that interaction of nuclear protein C with allele SNPG2 is sequence specific and suggest that nuclear protein C may be a member of the Egr transcription factor family. (The EMSA data shown in Fig. 5.9a lanes 2-6, represent preliminary competition data generated through part of the investigations into the sequence specificity of nuclear protein A binding to poly-G allele SNPG1. This data in Fig. 5.9a, lanes 2-6 is very similar to that already shown in Fig. 5.2b (page 177) and Fig. 5.3 (page 179) and therefore, will not be discussed further here).

The Egr family are zinc finger transcription factors (for review see O'Donovan et al, 1999). From the zinc titration experiment (Fig. 5.5, page 182) it is evident that nuclear protein C is sensitive to the presence of zinc in the EMSA binding reaction; notably forming a shifted protein-DNA complex of increasing intensity on probe SNPG2 as the concentration of zinc in the binding reaction is increased (Fig. 5.5, page 182, comparing lanes 3, 7 and 11). This suggests that nuclear protein C is a zinc finger transcription factor. Additionally, in these experiments it is evident that complex C is not observed on the SP1 consensus probe either in the absence or presence of zinc (Fig. 5.5 lanes 1, 5 and 9). This is in agreement with subsequent competition analysis; competition with the SP1 consensus sequence suggests that nuclear protein C is unlikely to be a member of the SP1 transcription factor family (Fig. 5.9b comparing lane 1 with 4 and 5). Instead, collectively these EMSA data suggest that OC-2 derived nuclear proteins C and A are distinct proteins.

To obtain further clarification as to whether nuclear protein C is a member of the Egr transcription factor family, the Egr consensus sequence was radiolabelled and incubated with OC-2 nuclear extracts (Fig. 5.9c). This approach showed that formation of shifted protein-DNA complexes on the Egr consensus probe is not analogous to those on probe SNPG2 (Fig. 5.9c). In particular, when the Egr consensus sequence is used as probe no band is visible at the site where complex C forms on probe SNPG2. Instead, on the Egr consensus probe a unique complex of slightly greater mobility, termed F, is observed lower down the gel (Fig. 5.9c, comparing lanes 2 and 3). This suggests that formation of complex C on probe SNPG2 may not be due to the binding of a member of the Egr transcription factor family. However, in the absence of further analysis such as supershift analysis with an anti-Egr antibody it is not possible to rule out that nuclear protein C is Egr.

In summary, these data suggest that nuclear protein C is a zinc finger transcription factor but more analysis than was possible within the scope of this project is needed to clarify whether nuclear protein C is a member of the Egr transcription factor family. However, the data is strongly suggestive that nuclear protein C is a different protein to nuclear protein A and not a member of the SP1 transcription factor family.

5.3.3.1 Effect of Zinc on binding of nuclear proteins A and C to common -3432poly-G alleles.

The data suggest that OC-2 derived nuclear proteins A and C are distinct proteins and in the absence of competitors exhibit a binding affinity for alleles SNPG1 and SNPG2, respectively (see Fig. 5.1, page 175, lanes 10, 11, 13 and 14; Fig. 5.5, page 182, lanes 2, 3, 6 and 7 and Fig. 5.16a lanes 1 and 2). This differential binding of nuclear proteins A and C to SNPG1 and SNPG2 probes is particularly evident under conditions of the zinc titration (Fig. 5.5) and suggests that the nature of the poly-G allele at position -3432 in the Brn-3c promoter modulates the binding of at least two, distinct nuclear proteins. To obtain further clarification this line of analysis was extended to the rest of the poly-G alleles and emphasis was placed on the binding of nuclear proteins A and C.

Radiolabelled probes for each of the common poly-G alleles were incubated with OC-2 nuclear extracts in the presence of increasing amounts of zinc (Fig. 5.10). This approach showed that in the presence of zinc nuclear protein A binding is evident on probe SNPG5 in addition to probe SNPG1 with little if any binding of nuclear protein C observed on these probes (Fig. 5.10 lanes 1 and 4). If the concentration of zinc

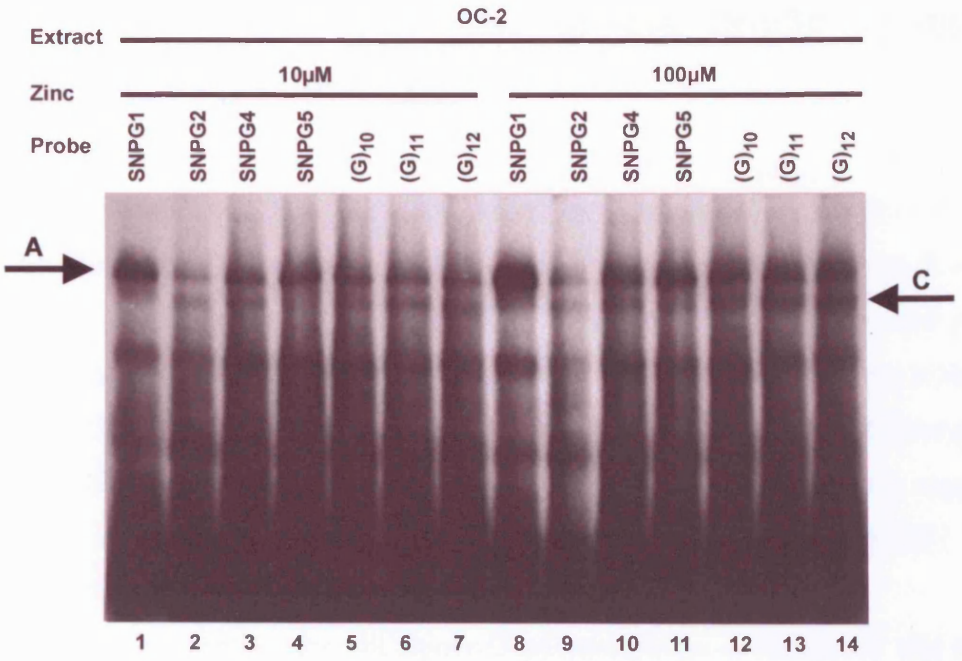


Figure 5.10 Effect of zinc on binding of nuclear proteins A and C to common -3432 poly-G alleles. Probes: SNPG1 (lanes 1 & 8), SNPG2 (lanes 2 & 9), SNPG4 (lanes 3 & 10), SNPG5 (lanes 4 & 11) $(G)_{10}$ (lanes 5 & 12), $(G)_{11}$ (lanes 6 & 13) or $(G)_{12}$ (lanes 7 &14) were incubated with 8μg OC-2 nuclear extract in the presence of 10μM zinc (lanes 1-7) or 100μM zinc (lanes 8-14). Arrows A and C denote the location of shifted nuclear protein-DNA complexes A and C, respectively.

in the binding reaction is increased this effect is augmented (Fig. 5.10 lanes 8 and 11). These results suggest that poly-G alleles SNPG1 and SNPG5 have a greater affinity for nuclear protein A compared to nuclear protein C. In contrast, in the presence of zinc, nuclear protein A binding in concert with nuclear protein C binding is enhanced at comparable amounts on poly-G probes SNPG2, SNPG4, $(G)_{10}$, $(G)_{11}$ and $(G)_{12}$ (Fig. 5.10 comparing lanes 3, 5, 6, 7, 10, 12, 13 and 14); this is particularly evident on probe $(G)_{12}$ (Fig. 5.10, lane 14). Hence, suggesting that the affinity of each of these poly-G alleles for nuclear proteins A and C is comparable in the presence of zinc. These data are in agreement with previous observations (Fig. 5.1, page 175; Fig. 5.5, page 182, and Fig. 5.9a) suggesting that the nature of the -3432poly-G allele plays an important role in determining the binding of two OC-2 derived nuclear proteins A and C to this region of the Brn-3c promoter. Nuclear protein C is discussed further in section 5.8.2 and for clarity an alternative exposure of the gel presented in figure 5.10 is shown in Appendix D.

5.4 Effect of the -3432poly-G polymorphism on basal Brn-3c promoter activity and activity of the Brn-3c promoter in response to exogenous SP1.

It has been established that the -3432poly-G polymorphism in the Brn-3c promoter modifies the binding affinity of at least two OC-2 derived nuclear proteins, A and C; the affinity of poly-G alleles for nuclear protein A being very similar to their respective abilities to bind purified recombinant SP1 protein. However, this does not mean that the -3432poly-G polymorphism is functional nor does it mean that the differential affinity in the -3432poly-G alleles for SP1 actually has a functional effect on regulation or expression of the Brn-3c gene. Indeed, sequence specific binding of SP1 does not necessarily correlate with promoter activity.

To assess whether the -3432poly-G polymorphism is functional and modulates transcriptional activity of Brn-3c a series of transient transfection experiments were performed in OC-2 cells. The pGL3-Basic vector (Promega) contains a modified firefly luciferase gene (*Fluc*) and vector backbone both designed to optimise expression in mammalian cells and to help reduce anomalous expression (see http://www.promega.com/tbs/tm033/tm033.pdf). The pGL3-Basic vector does not contain a eukaryotic promoter; transfecting cells with the empty pGL3-Basic vector (which lacks insertion of a putative regulatory sequence) produces only very low background luciferase expression driven by the vector backbone or cryptic regulatory elements present within the *Fluc* gene itself. Hence, efficient luciferase expression in cells transfected with pGL3-Basic vector results from sub-cloning a functional promoter fragment upstream of the *Fluc* gene. Thus, making the pGL3-Basic vector suitable for identifying putative regulatory sequences and for examining the effect of sequence variation on gene expression. The pGL3-Basic vector carrying a 3.6Kb fragment of the human Brn-3c promoter (bases -80bp to -3670bp with respect to the A of the ATG translation start as +1) was already cloned in our laboratory from the human PAC library RPCI1 (HGMP, MRC, U.K.) and previous experiments have shown that this fragment of the Brn-3c promoter is of sufficient length to drive *Fluc* expression in the inner ear sensory epithelial cell line, OC-2 (Jagutpal et al, unpublished data).

The region of the Brn-3c promoter upstream of the -3432poly-G polymorphism is extremely complex. Two repetitive sequence variants are present in close proximity with the -3432poly-G polymorphism: -3457$(GA)_{1\text{-}3}$ and -3495$(GT)_{15\text{-}21,24}$ and all

sequence variants at -3432, -3457 and -3495 appear in strong linkage disequilibrium. Thus, forming native haplotypes 5' of the -3432poly-G polymorphism (see Table 4.4, section 4.4.2, page 160 and Fig. 4.16, section 4.5, page 162). The complexity of this region of the Brn-3c promoter in the vicinity of the -3432poly-G polymorphism resulting from the high amount, the proximity and repetitive nature of variation identified limited the nature of functional analysis that could be performed.

For initial functional analysis -3432poly-G alleles SNPG1 and SNPG2 that were originally identified as part of the PCR-SSCP screen were focussed on. It is evident from EMSA analysis that -3432poly-G alleles SNPG1 and SNPG2 exhibit a high and low affinity respectively, for both OC-2 derived nuclear protein A and recombinant SP1 protein (see section 5.3) raising the possibility that these two alleles may have a differential functional effect(s) on activity of the Brn-3c promoter. To explore this possibility the firefly luciferase reporter vector pGL3-Basic (Promega) was used to generate two Brn-3c promoter-luciferase reporter gene constructs. Each construct contained 3.6Kb of the human Brn-3c promoter sub-cloned upstream of the firefly luciferase gene with either allele SNPG1 or allele SNPG2 at position -3432 in the Brn-3c promoter insert: p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$, respectively (a clear illustration of the variation present in the constructs is given in Fig. 5.11a, page 199; see method section 3.2.8.1 for details of construct formation). It was not possible to generate by site-directed mutagenesis or restriction digest sub-cloning Brn-3c promoter-luciferase reporter gene constructs that *only* varied in the nature of the poly-G allele at position -3432. The proximity and repetitive nature of the sequence variants at -3457 [$(GA)_{1-3}$], -3432 [poly-G polymorphism] and -3495 [$(GT)_{15-21,24}$] (see Fig. 4.16, section 4.5, page 162) made any attempts to alter appropriate Brn-3c promoter-luciferase reporter gene constructs by site-directed mutagenesis extremely problematic if not impossible. Site-directed mutagenesis is a PCR based technique and success is dependent on suitable DNA sequence flanking the target site for specific primer annealing; the highly repetitive nature of the -3432poly-G polymorphism (the target site) and surrounding promoter sequence prevented this line of analysis. In addition, no suitable restriction enzyme sites were found adjacent to the -3432poly-G polymorphism and prior to the -3457$(GA)_n$ repeat to facilitate sub-cloning of Brn-3c promoter-luciferase reporter gene constructs by restriction enzyme digest. Therefore, Brn-3c promoter-luciferase reporter gene constructs that differ in the nature of the poly-G allele (p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$) also differ in the length of the -3495 $(GT)_n$ repeat according to their native 5'-haplotypes (see Fig. 5.11a, page 199). For the

purpose of these experiments the GT repeat at -566 in the Brn-3c promoter [-566$(GT)_{17-23}$, see Fig. 4.16, section 4.5, page 162] was standardised at 23 repeats and this is reflected in the construct labelling: p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$ (the importance of the 5'haplotype in the Brn-3c promoter on regulation of the Brn-3c gene is explored further in Chapter 6 and is illustrated by analysis of the effect of the -566$(GT)_n$ variant on Brn-3c promoter activity).

To examine whether the nature of the poly-G allele at position -3432 in the Brn-3c promoter has an effect on basal expression levels of Brn-3c, Brn-3c promoter-luciferase reporter gene constructs p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$ (the experimental reporter gene constructs) were transfected into OC-2 cells along with the *renilla* luciferase expression vector pRL-SV40 (Promega) as an internal control (Fig. 5.11b). The *renilla* luciferase expression vector, pRL-SV40 contains the simian virus 40 (SV40) early enhancer / promoter region to provide strong constitutive expression from the renilla luciferase gene (*Rluc*) in transfected mammalian cells. By normalising the activity of *Fluc* from the experimental reporter gene constructs (p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$) to that of *Rluc* (the internal control reporter gene) intra- and inter- experimental variation resulting from differences in transfection efficiency and cell viability is minimised. Furthermore, because *Fluc* and *Rluc* activities are assayed sequentially within a single sample under conditions of the dual-luciferase® reporter assay system (Promega; see section 5.1.2) intra- and inter- experimental variation is reduced with a high degree of accuracy and variability manifesting from preparation of extracts for assay is eliminated. This line of analysis showed that basal activity between Brn-3c promoter-luciferase reporter gene constructs that differ in the nature of the poly-G allele at -3432 does not differ significantly; activity from both p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$ is approximately 3-fold higher than that from the empty pGL3-Basic vector alone (Fig. 5.11b; mean RLV: 272.04 ± 7.34 versus 281.80 ± 8.87, respectively). These results suggest that the nature of the poly-G allele at position -3432 and native 5'-haplotype does not appear to modulate basal transcriptional activity of Brn-3c, at least under these culture conditions. Indeed, the results imply that the differential affinity between poly-G alleles SNPG1 and SNPG2 for OC-2 derived nuclear protein A and / or C (section 5.3) has no affect on basal promoter activity of Brn-3c.

In the absence of observing an effect of the nature of the poly-G allele on basal promoter activity of Brn-3c and in order to assess whether the high affinity SP1 binding site, allele SNPG1 or the low affinity SP1 binding site, allele SNPG2 has a functional

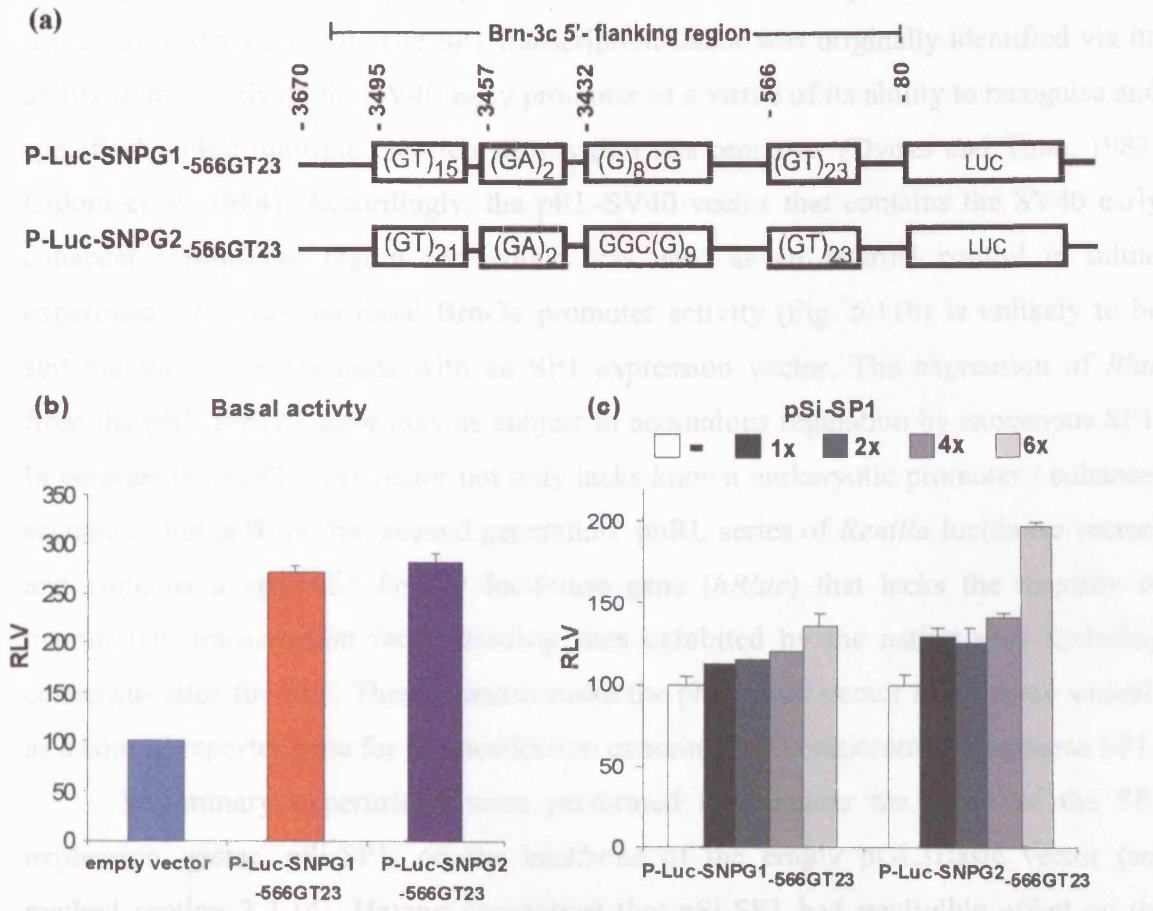


Figure 5.11 Effect of the -3432 poly-G polymorphism on basal Brn-3c promoter activity and activity of the Brn-3c promoter in response to exogenous SP1. (a) A schematic diagram (not to scale) to illustrate the 5' haplotype of the 3.6Kb Brn-3c promoter-luciferase reporter gene constructs that contain either allele SNPG1 or allele SNPG2 at position -3432. **(b)** OC-2 cells were transfected with 500ng of either p-Luc-SNPG1$_{566GT23}$, p-Luc-SNPG2$_{566GT23}$ or the empty pGL3-Basic vector. The renilla expression vector pRL-SV40 (100ng) was used as an internal control to normalise for differences in transfection efficiency. **(C)** OC-2 cells were co-transfected with 500ng of either p-Luc-SNPG1$_{566GT23}$ or p-Luc-SNPG2$_{566GT23}$ in the presence of increasing amounts (0, 0.5, 1, 2 & 3μg) of SP1 expression vector (pSi-SP1). The empty pSi expression vector was used to standardise DNA concentrations between wells and the renilla expression vector phRL-null (10ng) was used as an internal control to normalise for differences in transfection efficency. Error bars on charts represent the standard error of the mean. RLV: relative luciferase value. Luciferase activities shown on charts are expressed: (b) relative to the empty vector or (C): relative to each construct in the absence of stimulation by assigning a RLVof 100. Each experiment was performed in triplicate and repeated at least twice with two different preparations of DNA.

effect on Brn-3c promoter activity, activity of p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$ in response to exogenous SP1 from an SP1 expression vector was examined (Fig.5.11c). The mammalian expression vector pEVR2-Sp1 containing the complete human cDNA for SP1 has been described previously (Hagen et al, 1992) and was kindly provided by Professor Guntram Suske, University of Marburg, Germany. Using vector pEVR2-Sp1 the cDNA for SP1 was sub-cloned into the pSi mammalian expression vector (Promega) for constitutive expression in the OC-2 cell line (see method section 3.2.8.2). In these experiments the *renilla* luciferase expression vector phRL-null

(Promega) that lacks eurkaryotic promoter and /or enhancer sequences was selected for use as an internal control. The SP1 transcription factor was originally identified via its ability to transactivate the SV40 early promoter as a virtue of its ability to recognise and specifically bind multiple GC-rich sites within this promoter (Dynan and Tjian, 1983; Gidoni et al, 1984). Accordingly, the pRL-SV40 vector that contains the SV40 early enhancer / promoter region and which was used as an internal control in initial experiments to examine basal Brn-3c promoter activity (Fig. 5.11b) is unlikely to be suitable for co-transfections with an SP1 expression vector. The expression of *Rluc* from the pRL-SV40 vector may be subject to anomalous regulation by exogenous SP1. In contrast the phRL-null vector not only lacks known eurkaryotic promoter / enhancer sequences but is from the 'second generation' phRL series of *Renilla* luciferase vectors and contains a synthetic *Renilla* luciferase gene (*hRluc*) that lacks the majority of mammalian transcription factor binding sites exhibited by the native gene including consensus sites for SP1. These features make the phRL-null vector much more suitable as a control reporter gene for co-transfection experiments incorporating exogenous SP1.

Preliminary experiments were performed to examine the effect of the SP1 expression vector, pSi-SP1, on the backbone of the empty pGL3Basic vector (see method section 3.2.14). Having ascertained that pSi-SP1 had negligible effect on the backbone of the empty pGL3Basic vector increasing amounts of pSi-SP1 were co-transfected into OC-2 cells alongside Brn-3c promoter-luciferase reporter gene constructs that carried either the high affinity SP1 binding allele SNPG1 (p-Luc-SNPG1$_{-566GT23}$) or the low affinity SP1 binding allele SNPG2 (p-Luc-SNPG2$_{-566GT2}$) at position -3432 (see, Fig. 5.11c). In each titration experiment the empty pSi expression vector (Promega) was used to standardise DNA concentrations between wells to minimise the possible affect of inter-well variation in DNA concentration on transfection efficiency. This line of analysis showed that SP1 has little effect on the Brn-3c promoter construct carrying allele SNPG1; two, four and six-fold excess of pSi-SP1 vector only increases Brn-3c promoter activity by 14, 19 and 35%, respectively (Fig. 5.11c). In contrast a somewhat greater, although not significantly different response is observed if allele SNPG2 is present; two, four and six-fold excess of pSi-SP1 vector increases expression from p-Luc-SNPG2$_{-566GT23}$ by 25, 41 and 96%, respectively (Fig. 5.11c; these results are not significantly different when tested by paired t-test analysis). These results show that neither Brn-3c promoter-luciferase reporter gene construct that carries allele SNPG1 or allele SNPG2 is strongly transactivated by exogenous SP1. Hence, suggesting that SP1 has little effect on

regulation of Brn-3c promoter activity. This is despite the finding that EMSA data show allele SNPG1 to contain a high affinity SP1 binding site (section 5.3.2). It is possible that the SP1 binding site at -3432 in the Brn-3c promoter, which is modulated by the nature of the poly-G alleles, does not play an important role in Brn-3c regulation. Alternatively, it is possible that endogenous SP1 expression in OC-2 cells may not be limiting under these experimental conditions and hence adding a source of exogenous SP1 into these cells does not influence Brn-3c promoter activity. Indeed, SP1 is a housekeeping gene and expression is thought to be ubiquitous (for reviews see Philipsen and Suske, 1999; Kaczynski et al, 2003; Zhao and Meng, 2005). In the mouse, levels of SP1 vary greatly during embryonal development; depending on the cell type and stage of development the levels of SP1 mRNA vary at least 100-fold and this is reflected at the protein level (Saffer et al, 1991). However, it has not been ascertained whether the OC-2 cell line that is derived from inner ear sensory epithelia in the H-2Kb-*ts*A58 Immortomouse at E13 immediately before hair cells undergo their final mitosis express SP1.

In light of these findings western immunoblot analysis was performed on OC-2 nuclear extracts to assess whether the OC-2 cell line expresses SP1 (Fig. 5.12). The mouse monoclonal anti-SP1 antibody [SP1 (1C6) X, Santa Cruz Biotechnology, Inc.] used in supershift analysis (section 5.3.1.3) was used at a dilution of 1:10,000. In mammalian cells SP1 exists in two isoforms of 95 and 106Kda. Western immunoblot analysis of OC-2 nuclear extracts revealed a clear band immediately below the standard Rainbow™ coloured protein molecular weight marker band of 97.4 kDa (Fig. 5.12). This is in agreement with the 95kDa isoform for SP1 and confirms SP1 is expressed in the OC-2 cell line. Although, attempts were not made to quantify the amount of SP1 protein expressed it is clear that the OC-2 cell line expresses SP1 and therefore, it is possible that SP1 is not limiting in OC-2 cells under the culture conditions used.

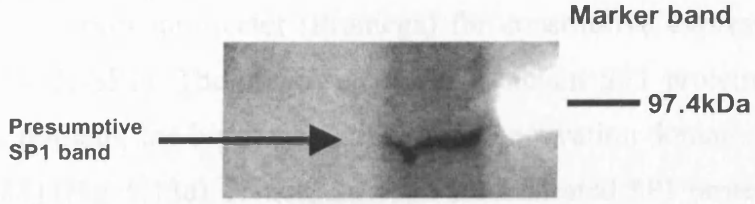


Figure 5.12 Western immunoblot of OC-2 nuclear extracts with anti-SP1 antibody. 24μg OC-2 nuclear extract was resolved via 8% SDS-polyacrylamide gel electrophoresis, transferred to a Hybond ™-C nitrocellulose membrane and probed with primary antibody (diluted 1:10,000): mouse monoclonal anti-SP1 antibody [SP1 (1C6) X, Santa Cruz Biotechnology Inc.]. Secondary antibody HRP-conjugated goat anti-mouse antibody (DAKO) was used at a 1:1000 dilution. Arrow denotes the presumptive SP1 band. Position of the 97.4kDa Rainbow™ protein molecular weight marker band is indicated.

Accordingly, it is feasible that the presence of endogenous SP1 in OC-2 cells could have confounded analysis using an SP1 expression vector and masked any effect of exogenous SP1 on Brn-3c promoter activity.

5.4.1 A dominant negative approach to assess whether the differential affinity for SP1 at the -3432 polymorphic site has a functional affect on Brn-3c promoter activity.

It is possible that endogenous SP1 levels are not limiting in OC-2 cells and that this is responsible for the limited effect observed when an SP1 expression vector is co-transfected with Brn-3c promoter-luciferase reporter gene constructs containing either the high affinity SP1 binding allele, SNPG1 (p-Luc-SNPG1$_{-566GT23}$) or the low affinity SP1 binding allele, SNPG2 (p-Luc-SNPG2$_{-566GT23}$) (Fig. 5.11c, page 199). In agreement with this suggestion Western immunoblot analysis confirms that the OC-2 cell line expresses SP1 (Fig. 5.12). Therefore, a dominant negative approach using a truncated form of SP1 was undertaken in an attempt to assess the affect of the differential affinity for SP1 at the -3432 polymorphic site when SP1 levels are limiting.

The dominant negative SP1 expression vector pEBGN-SP1 has been described previously (Petersohn and Thiel, 1996) and has been successfully used in transient transfection assays to establish the role of SP1 in gene regulation (Lietz et al, 2003; Al-Sarraj et al, 2005). The pEBGN-SP1 vector encodes a fusion protein of glutathione S-transferase (GST) linked to the DNA binding domain (DBD) of SP1 under the control of the human polypeptide chain elongation factor 1α promoter. It is not known whether this promoter is active in OC-2 cells. In order to use this truncated form of SP1 in OC-2 cells PCR primers were designed to amplify a fragment from the pEBGN-SP1 vector that contained the cDNA for the SV40 nuclear localisation signal (NLS) linked to the DBD of SP1 (see method section 3.2.8.2) and this fragment was subcloned into the pSi mammalian expression vector (Promega) for constitutive expression in the OC-2 cell line (pSi-DBD-SP1). The dominant negative mutant SP1 protein lacks the N-terminal activation domain; the bipartite glutamine rich activation domain is absent (Courey and Tjian, 1988) (Fig. 5.13a). Hence, although the truncated SP1 protein is able to recognise and bind to its cognate sites it is unable to activate gene expression. This means that exogenous dominant negative SP1 protein has the ability to compete with endogenous SP1 for SP1 DNA-binding elements and once bound to render these sites inaccessible to endogenous SP1 and block activation.

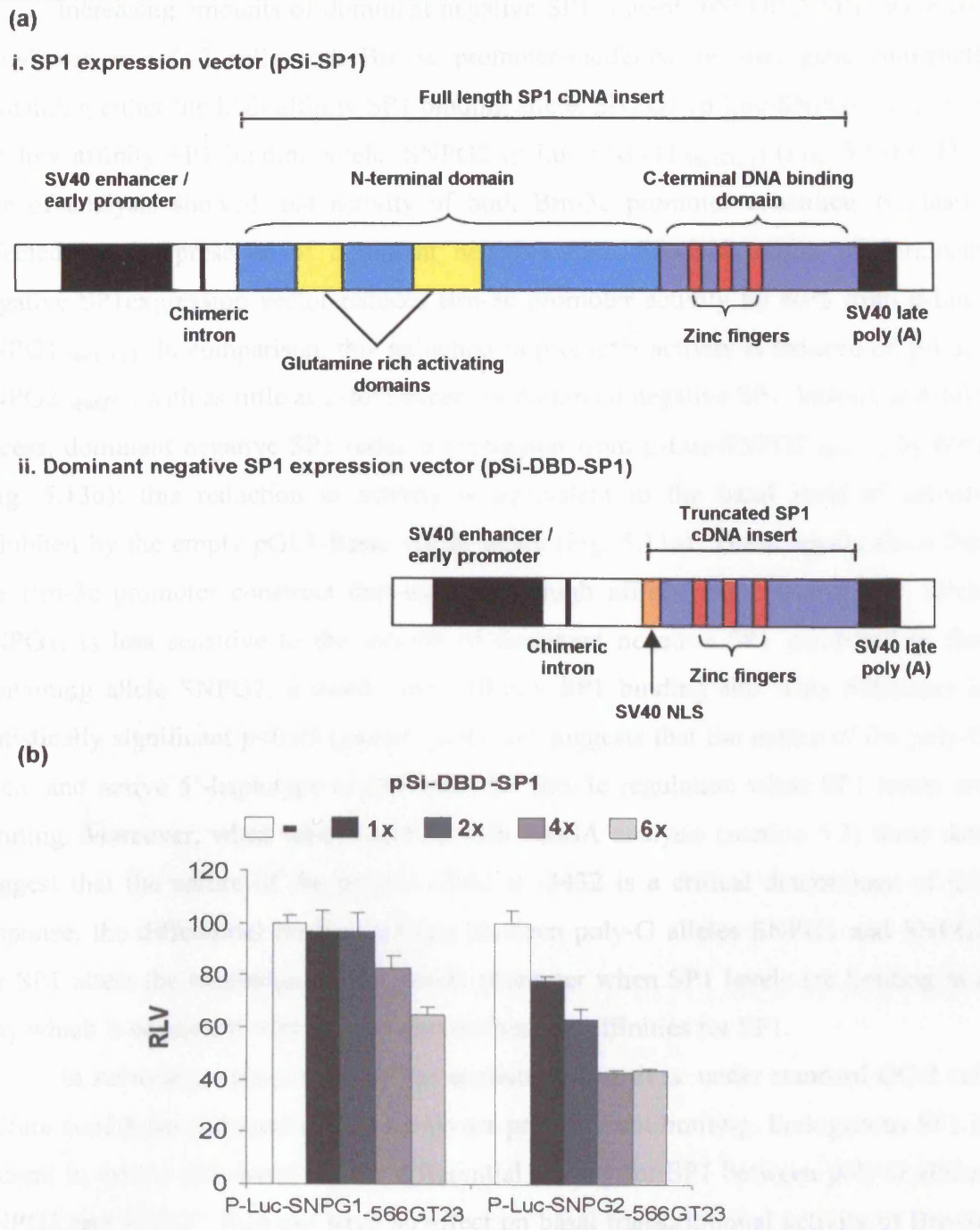


Figure 5.13 A dominant negative approach to assess whether the differential affinity for SP1 at the -3432 polymorphic site has a functional affect on Brn-3c promoter activity. (a i & ii) A schematic diagram (not to scale) to illustrate dominant negative SP1 expression vector compared to SP1 expression vector. Features of the pSi expression vector backbone (Promega) are shown as black shaded boxes. **i.** SP1 expression vector illustrating the N-terminal glutamine rich activating domains and C-terminal zinc finger DNA binding domain of SP1 (Hagen et al, 1992). **ii.** Dominant negative SP1 lacks the N-terminal portion of the protein (Petersohn and Thiel, 1996). The SV40 nuclear localisation signal (NLS) is 5' to the zinc finger DNA binding domain (orange shaded box). **(b)** OC-2 cells were co-transfected with 500ng of either p-Luc-SNPG1$_{566GT23}$ or p-Luc-SNPG2$_{566GT23}$ in the presence of increasing amounts (0, 0.5, 1, 2 & 3μg) of dominant negative SP1 expression vector (pSi-DBD-SP1). The empty pSi expression vector was used to standardise DNA concentrations between wells and the renilla expression vector phRL-null (10ng) was used as an internal control to normalise for differences in transfection efficency. Error bars on charts represent the standard error of the mean. Luciferase activities are expressed relative to each construct in the absence of stimulation by assigning a relative value of 100 (RLV). The experiment was performed in triplicate and repeated at least twice with two different preparations of DNA. The difference in activity between p-Luc-SNPG1$_{566GT23}$ and p-Luc-SNPG2$_{566GT23}$ constructs is significantly different at $p<0.05$ (paired t-test; d.f. = 4).

Increasing amounts of dominant negative SP1 mutant (pSi-DBD-SP1) were co-transfected in OC-2 cells with Brn-3c promoter-luciferase reporter gene constructs containing either the high affinity SP1 binding allele, SNPG1 (p-Luc-SNPG1$_{-566GT23}$) or the low affinity SP1 binding allele, SNPG2 (p-Luc-SNPG2$_{-566GT23}$) (Fig. 5.13b). This line of analysis showed that activity of both Brn-3c promoter constructs is clearly affected by the presence of dominant negative SP1. Six-fold excess of dominant negative SP1expression vector reduces Brn-3c promoter activity by 40% from p-Luc-SNPG1$_{-566GT23}$. In comparison, this reduction in promoter activity is induced on p-Luc-SNPG2$_{-566GT23}$ with as little as 2-fold excess of dominant negative SP1. Indeed, at 4-fold excess, dominant negative SP1 reduces expression from p-Luc-SNPG2$_{-566GT23}$ by 60% (Fig. 5.13b); this reduction in activity is equivalent to the basal level of activity exhibited by the empty pGL3-Basic vector alone (Fig. 5.11a). These results show that the Brn-3c promoter construct that carries the high affinity SP1 binding site, allele SNPG1, is less sensitive to the activity of dominant negative SP1 compared to that containing allele SNPG2, a much lower affinity SP1 binding site. This difference is statistically significant $p<0.05$ (paired t-test) and suggests that the nature of the poly-G allele and native 5'-haplotype is important for Brn-3c regulation when SP1 levels are limiting. Moreover, when taken together with EMSA analysis (section 5.3) these data suggest that the nature of the poly-G allele at -3432 is a critical determinant of this response; the differential binding affinity between poly-G alleles SNPG1 and SNPG2 for SP1 alters the regulation of the Brn-3c promoter when SP1 levels are limiting in a way which is consistent with their respective binding affinities for SP1.

In summary, a possible model to explain this result is: under standard OC-2 cell culture conditions endogenous SP1 levels are probably not limiting. Endogenous SP1 is present in excess and therefore, the differential affinity for SP1 between poly-G alleles SNPG1 and SNPG2 does not have an effect on basal transcriptional activity of Brn-3c (Fig. 5.11b). Consequently, adding exogenous SP1 in the form of the pSi-SP1 expression vector has very little effect on Brn-3c promoter activity (Fig. 5.11c). However, if the culture conditions are changed such that SP1 levels are limiting by co-transfection of dominant negative SP1 mutant (pSi-SP1-DBD) the differential affinity for SP1 between poly-G alleles SNPG1 and SNPG2 has an affect on Brn-3c promoter activity (Fig. 5.13b). The high affinity SP1 binding site, allele SNPG1 is presumably able to retain endogenous SP1, or at least for longer periods compared to allele SNPG2 and therefore is not readily susceptible to the presence of excess dominant negative SP1 (Fig. 5.13b). In contrast, under these culture conditions the low affinity SP1 binding

site, allele SNPG2 is probably less able to retain endogenous SP1 binding and hence, readily succumbs to the presence of excess dominant negative SP1 (Fig. 5.13b). This putative model is summarised in Fig. 5.14.

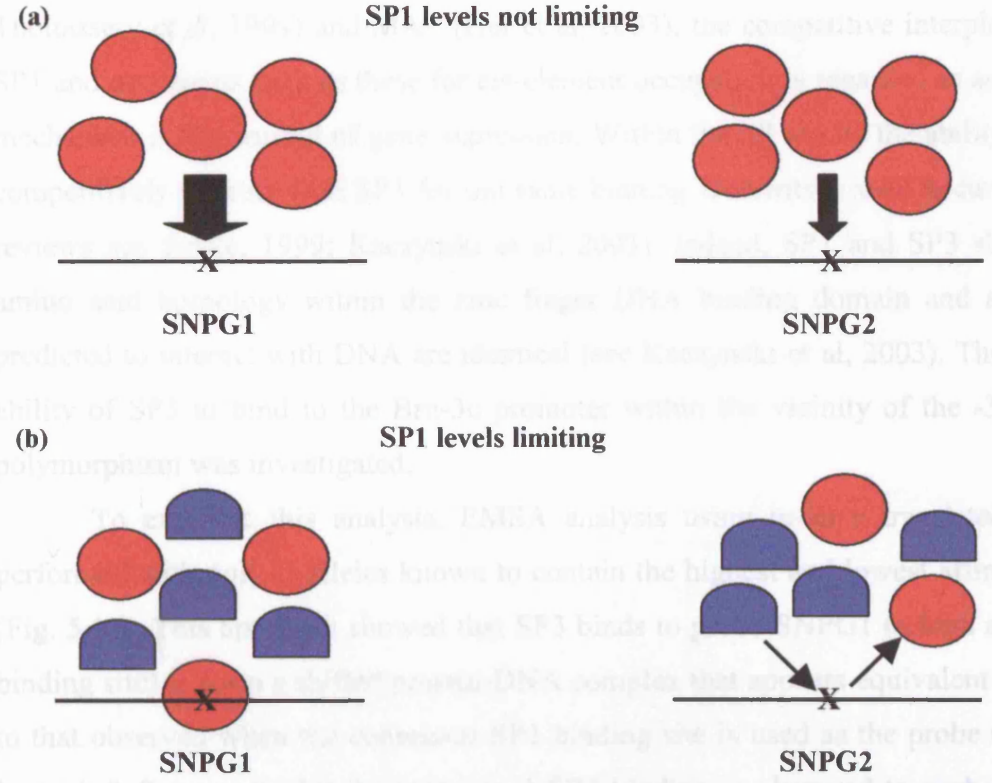


Figure 5.14 A putative model to illustrate how the differential binding affinity between poly-G alleles SNPG1 and SNPG2 for SP1 alters the regulation of the Brn-3c promoter when SP1 levels are limiting. **(a)** When SP1 levels are not limiting the differential affinity for SP1 between poly-G alleles SNPG1 and SNPG2 (illustrated by thick and thin black shaded arrows, respectively) does not have an affect on Brn-3c promoter activity; the high affinity SP1 binding allele, SNPG1, readily binds SP1 (shown as a red shaded circle) and the low affinity SP1 binding allele, SNPG2, can bind SP1 albeit with lower affinity. **(b)** When SP1 levels are limiting, the high affinity SP1 binding allele, SNPG1, retains endogenous SP1, or at least for longer periods compared to allele SNPG2 and is not readily susceptible to the presence of excess dominant negative SP1 (shown as a blue shaded arc). In contrast, the low affinity SP1 binding allele, SNPG2, is less able to retain endogenous SP1 binding and readily succumbs to the presence of excess dominant negative SP1. X denotes the -3432 poly-G polymorphic site in the Brn-3c promoter.

5.5 The -3432poly-G polymorphism modulates the binding of SP3.

SP1 has been shown to competitively interact with other transcription factors for binding to the same cis-acting motifs including Egr-1 (Ackerman et al, 1991; Thottassery et al, 1999) and MAZ (Her et al, 2003); the competitive interplay between SP1 and opponents such as these for cis-element occupation is regarded as an important mechanism in the control of gene expression. Within the SP family the ability of SP1 to competitively interact with SP3 for the same binding elements is well documented (for reviews see Suske, 1999; Kaczynski et al, 2003). Indeed, SP1 and SP3 share 95.1% amino acid homology within the zinc finger DNA binding domain and all residues predicted to interact with DNA are identical (see Kaczynski et al, 2003). Therefore, the ability of SP3 to bind to the Brn-3c promoter within the vicinity of the -3432poly-G polymorphism was investigated.

To expedite this analysis, EMSA analysis using *in-vitro* translated SP3 was performed with poly-G alleles known to contain the highest and lowest affinity for SP1 (Fig. 5.15). This approach showed that SP3 binds to probe SNPG1 (a high affinity SP1 binding site) to form a shifted protein-DNA complex that appears equivalent in intensity to that observed when the consensus SP1 binding site is used as the probe (Fig. 5.15a, lanes 1 & 2 respectively). In contrast no SP3 binding is observed to probe SNPG2 (a low affinity SP1 binding site) (Fig. 5.15a lane 3), presumably because allele SNPG2 has a much-reduced affinity for SP3. Competition analysis with 25-fold excess of the unlabelled SNPG1 probe induces a moderate reduction in the intensity of the shifted complex formed by the binding of *in-vitro* translated SP3 to the SP1 consensus sequence, with the total disappearance of this shifted complex evident at 200-fold competition (Fig. 5.15b, lanes 1, 3 and 6). In comparison, the unlabeled SNPG2 probe does not appear to induce any significant reduction of the shifted complex (Fig. 5.15b, lanes 1, 4 and 7). Taken together, these results confirm that the high affinity SP1 binding allele, SNPG1 has a much greater affinity for SP3 compared to the low affinity SP1 binding allele, SNPG2. Similar analysis was performed using *in-vitro* translated SP1 and this was in agreement with the data obtained using OC-2 nuclear extracts and recombinant SP1 protein (section 5.3).

Having established that allele SNPG1, a high affinity SP1 binding site, is also capable of binding SP3, the ability of SP1 and SP3 to competitively interact at this site was examined. The SNPG1 probe was incubated with increasing amounts of *in-vitro*

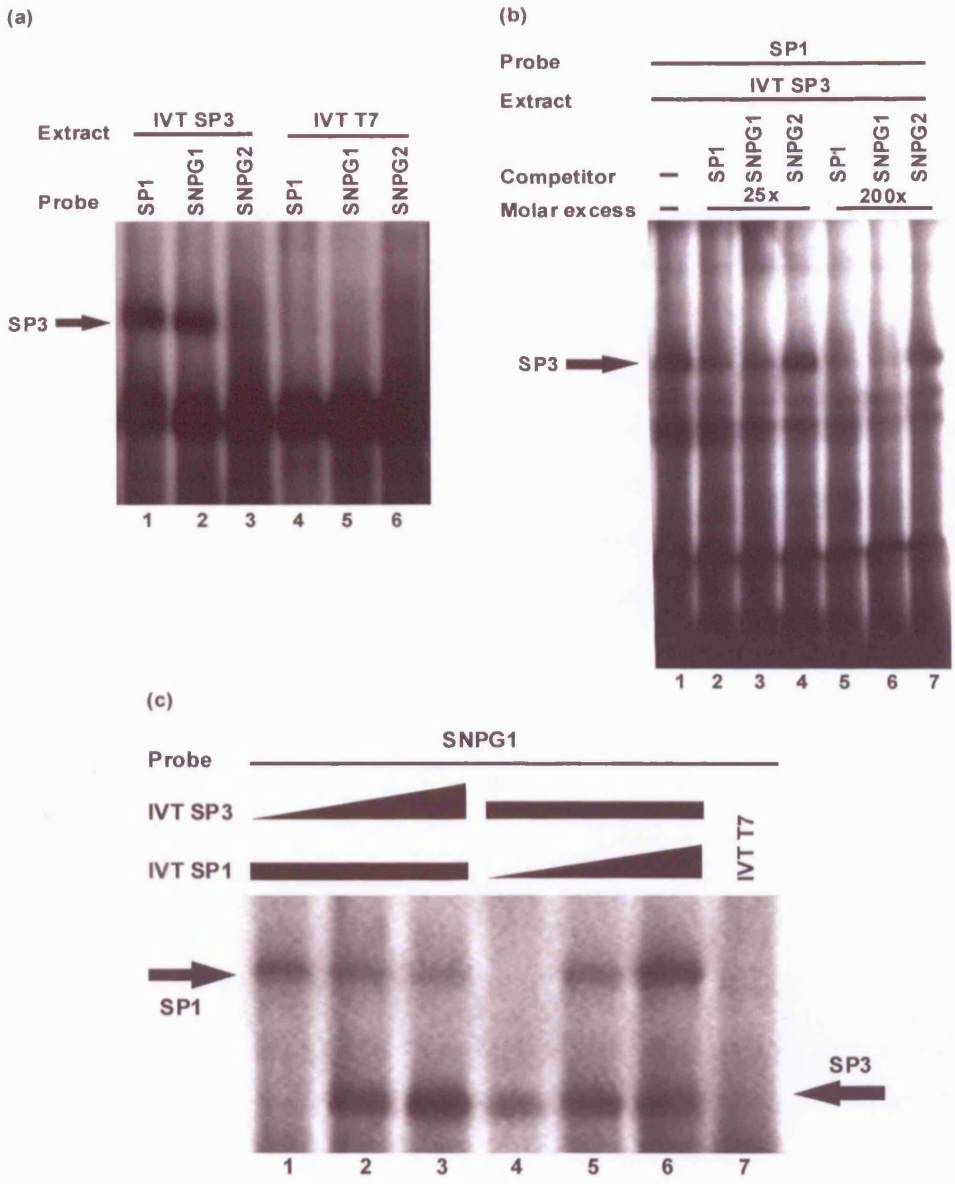


Figure 5.15 Effect of -3432 poly-G polymorphism on binding of SP3. (a) The labelled SP1 consensus sequence (lanes 1 & 4), SNPG1 probe (lanes 2 & 5) or SNPG2 probe (lanes 3 & 6) were incubated in the presence of 5µl SP3 *in-vitro* translate (lanes 1-3) or 5µl T7 *in-vitro* translate as a negative control for endogenous protein binding (lanes 4-6). The T7 *in-vitro* translate contains a luciferase encoding plasmid under the control of the T7 RNA polymerase promoter (Promega). The arrow denotes the location of the shifted protein-DNA complex formed by the binding of IVT SP3 to the labelled probes. **(b)** The labelled SP1 consensus sequence was incubated in the presence of 5µl SP3 *in-vitro* translate in the absence of competitor (lane 1) or in the presence of a molar excess of unlabelled SP1 consensus sequence (lanes 2 & 5), SNPG1 (lanes 3 & 6) or SNPG2 (lanes 4 & 7). **(c)** The SNPG1 probe was incubated in the presence of 3.5µl SP1 *in-vitro* translate (lanes 1-3) in the presence of increasing amounts of SP3 *in-vitro* translate: 1µl (lane 2) and 2µl (lane 3); 1µl SP3 *in-vitro* translate (lanes 4-6) in the presence of increasing amounts of SP1 *in-vitro* translate: 3.5µl (lane 5) and 7µl (lane 6) or 8µl T7 *in-vitro* translate as a negative control for endogenous protein binding. Figure 5.15c is reproduced in full in Appendix E.

translated SP3 as the amount of *in-vitro* translated SP1 in the binding reaction was held constant (Fig. 5.15c). This line of analysis showed that the complex formed by the binding of *in-vitro* translated SP1 to probe SNPG1 progressively diminishes in intensity

as the complex formed by the binding of *in-vitro* translated SP3 to probe SNPG1 intensifies. Hence suggesting that SP3 is able to competitively displace SP1 from the high affinity SP1 binding site on allele SNPG1. When the reverse experiment was performed and the amount of *in-vitro* translated SP3 in the binding reaction was held constant as the amount of *in-vitro* translated SP1 was increased (Fig. 5.15c) the complex formed by the binding of *in-vitro* translated SP3 to probe SNPG1 did not appear to diminish in intensity although the complex for *in-vitro* translated SP1 clearly intensified. These results suggest that SP3 is somewhat effective at displacing SP1 from probe SNPG1 but clearly, SP1 shows little ability to displace SP3, at least under these assay conditions. In light of these preliminary observations it is possible that SP1 and SP3 competitively interact for SNPG1 allele binding. However, based on the limited preliminary data available it is not possible to ascertain whether this is the case.

5.6 Effect of the differential affinity for SP3 at the -3432 polymorphic site on Brn-3c promoter activity.

It is widely recognised that SP1 functions as a transactivator of transcription, and SP3 as a repressor, able to repress SP1-mediated transactivation (Hagen et al, 1994; for review see Suske, 1999). EMSA analysis using *in-vitro* translated SP3 shows that SP3 can bind with high affinity to allele SNPG1, a high affinity SP1 binding site and similarly the low affinity SP1 allele, SNPG2, shows a much-reduced affinity for SP3 (section 5.5). To assess whether the differential affinity for SP3 at the -3432 polymorphic site has a functional affect on regulation of the Brn-3c gene a titration experiment with an expression vector for SP3 was performed in OC-2 cells.

The *Drosophilia* expression vector pPac-USp3 containing the complete human SP3 cDNA has been described previously (Dennig et al, 1996) and was kindly provided by Professor Guntram Suske, University of Marburg, Germany. Using vector pPac-USp3 the cDNA for SP3 was sub-cloned into the pSi mammalian expression vector (Promega) for constitutive expression in the OC-2 cell line (see method section 3.2.8.2). Preliminary experiments were performed to examine the effect of the SP3 expression vector, pSi-SP3, on the backbone of the empty pGL3Basic vector (see method section 3.2.14). Having ascertained that pSi-SP3 had negligible effect on the backbone of the empty pGL3Basic vector increasing amounts of SP3 expression vector (pSi-SP3) were co-transfected with Brn-3c promoter-luciferase reporter gene constructs containing either the high affinity SP binding site, allele SNPG1 (p-Luc-SNPG1.$_{-566GT23}$) or the low affinity SP binding site, allele SNPG2 (p-Luc-SNPG2.$_{-566GT23}$) in the presence of phRL-null vector as an internal control (Fig. 5.16). This approach showed that SP3 has only a moderate effect on the Brn-3c promoter construct carrying allele SNPG1; two, four and six-fold excess of pSi-SP3 vector reduced Brn-3c promoter activity by 5, 21 and 34% respectively (Fig. 5.16). In contrast no appreciable reduction in activity is observed if the Brn-3c promoter construct contains allele SNPG2, even at six-fold excess pSi-SP3 vector expression is not significantly reduced from p-Luc-SNPG2.$_{-566GT23}$ (Fig. 5.16). These observations are consistent with the respective binding affinities of alleles SNPG1 and SNPG2 for SP3 in EMSA analysis (see section 5.3, Fig. 5.15a and b) but are not significantly different when tested by paired t-test analysis. Taken together, this data suggests that the differential affinity for SP3 at the -3432 polymorphic site has some effect, albeit subtle, on regulation of Brn-3c promoter activity although not at statistically significantly different levels (paired t-test analysis). Expression of SP3

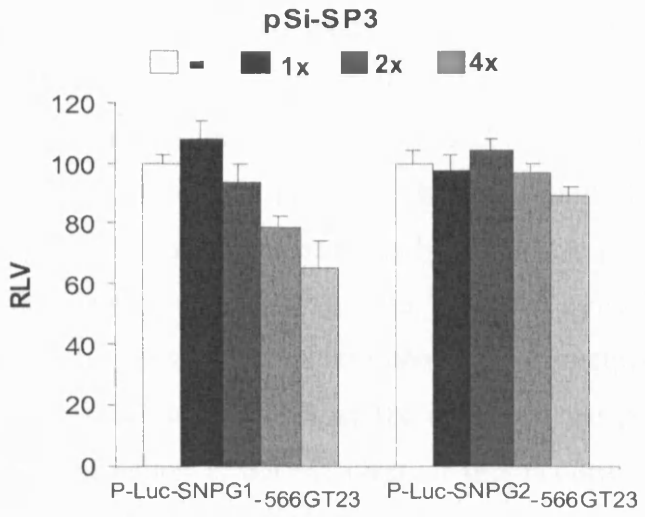


Figure 5.16 Effect of the -3432poly-G polymorphism on activity of the Brn-3c promoter in response to exogenous SP3. OC-2 cells were co-transfected with 500ng of either p-Luc-SNPG1$_{566GT23}$ or p-Luc-SNPG2$_{566GT23}$ in the presence of increasing amounts (0, 0.5, 1, 2 & 3μg) of SP3 expression vector (pSi-SP3). The empty pSi expression vector was used to standardise DNA concentrations between wells and the renilla expression vector phRL-null (10ng) was used as an internal control to normalise for differences in transfection efficency. Error bars on charts represent the standard error of the mean. Luciferase activities are expressed relative to each construct in the absence of stimulation by assigning a relative value of 100 (RLV). The experiment was performed in triplicate and repeated at least twice with two different preparations of DNA.

similar to that of its sister factor, SP1, is thought to be ubiquitous (for reviews see Kaczynski et al, 2003; Zhao and Meng, 2005). Accordingly, it is possible that the subtle differences in Brn-3c promoter activity as observed are due to endogenous expression of SP3 in OC-2 cells.

Interestingly, a zinc titration in the EMSA binding assay revealed enhanced formation of a nuclear protein-DNA complex, designated D, on probes SNPG1 and SNPG2 as the concentration of zinc in the binding reaction was increased; this complex is arguably more prominent on probe SNPG1 compared to probe SNPG2 (see Fig. 5.5 comparing lanes 2 and 3 with 6 and 7, and 10 and 11, page 182). These findings suggests the presence of a nuclear protein(s) (hereafter referred to as nuclear protein D) that may be a zinc finger transcription factor and which may have a greater affinity for allele SNPG1 compared to allele SNPG2. It was not possible to investigate the identity of nuclear protein D further within the scope of this project but from this limited data it is tempting to speculate that it may be SP3.

Depending on the promoter and cellular context, SP1-mediated promoter activity can be attenuated by SP3 (Hagen et al, 1994; for review see Suske, 1999). Indeed, for some SP1-responsive genes the nuclear SP1:SP3 ratio appears to be important in regulating gene expression (Hoppe and Francone, 1998; Hata et al, 1998).

From the limited competitive interaction studies presented in this thesis using *in-vitro* translated SP1 and SP3 (Fig. 5.15c, page 207) it was not possible to confirm that SP1 and SP3 competitively interact for SNPG1 allele binding. However, it would be interesting to investigate this possibility further and to extend this line of analysis to the rest of the -3432poly-G alleles in particular the low affinity SP binding allele SNPG2. At present it is clear that the nature of the poly-G allele at position -3432 in the Brn-3c promoter is a determinant of SP1 (Fig. 5.8a, page 189) and SP3 binding (Fig. 5.15b, page 207) but whether this translates into competitive interplay between SP1 and SP3 for binding site occupation and whether the nature of the poly-G allele modifies the dynamics of this interaction is not yet clear. If this occurred it would be tempting to speculate that the nuclear SP1:SP3 ratio is at least one mechanism whereby the -3432poly-G polymorphism modulates expression of the Brn-3c gene. To explore this notion, extended competitive interaction studies would need to be performed and supported with transient transfection assays using SP1 and SP3 expression vectors. In this regard *Drosophila* Schneider's SL-2 cells may prove useful (Schneider, 1972). Unlike the sensory epithelial OC-2 cell line derived from the inner ear of Immortomouse, SL-2 cells are an embryonic cell line derived from *Drosophila* and hence, do not resemble the physiological context of hair cells. However, SL-2 cells are devoid of endogenously expressed SP proteins and have been widely used to investigate the affect of SP1 and / or SP3 on the control of human gene expression (Courey and Tjian, 1988; Hoppe and Francone, 1998; Suske, 2000; Koutsodontis et al, 2001).

5.7 Effect of the -3432poly-G polymorphism in the context of a heterologous promoter.

The transient transfection assays designed to investigate the effect of the -3432poly-G polymorphism on Brn-3c promoter activity have so far focussed on using 3.6Kb Brn-3c promoter fragments containing conserved haplotypes 5' of the -3432 poly-G polymorphism (the -3495$(GT)_n$ repeat and -3457$(GA)_n$ repeat appear in strong linkage disequilibrium with the -3432 poly-G polymorphism; see Table 4.4, section 4.4.2, page 160). Consequently it is important to remember that Brn-3c promoter constructs that differ in the nature of the poly-G allele (p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$) also differ in the length of the -3495$(GT)_n$ repeat according to their native 5' haplotype (see Fig. 5.11a for haplotype of Brn-3c promoter constructs, page 199). The precise contribution, if any, of the -3495$(GT)_n$ repeat on basal promoter activity of Brn-3c (Fig. 5.11b, page 199) or on activity of Brn-3c promoter constructs when co-transfected with an expression vector for SP1 (Fig. 5.11c, page 199), dominant negative SP1 (Fig. 5.13b, page 203) or SP3 (Fig. 5.16b, page 210) cannot be delineated from these data. Indeed, it is impossible to exclude that the -3495$(GT)_n$ repeat has an affect on the significantly different activity of the Brn-3c promoter when constructs p-Luc-SNPG1$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT23}$ are co-transfected with an expression vector for dominant negative SP1 (Fig. 5.13b, page 203). The repetitive nature of the sequence variants and their close proximity meant it was not possible to generate constructs that only varied at the -3432 poly-G position. However, the EMSA data is very strongly suggestive that the region spanning the -3432poly-G polymorphism is responsible for the altered response of the Brn-3c promoter to SP1 and SP3 (see section 5.3 and 5.5). EMSA analysis using different lengths of the -3495$(GT)_n$ repeat as probe to assess whether SP1 and SP3 can bind to this sequence and whether binding is modified by repeat length may have been informative. However, there is very little evidence that transcription factors can bind dinucleotide repeats compared to much evidence they can bind tri-, tetra-, and penta-nucleotide repeat motifs (Albanese et al, 2001; Contente et al, 2002; Iglesias et al, 2004). Certainly a search of the literature revealed no evidence that SP1 or SP3 can bind GT dinucleotide repeats. Therefore, in an attempt to obtain further clarification of this issue the functional significance of the -3432poly-G polymorphism was examined in the context of the SV40 heterologous promoter in contrast to the native Brn-3c promoter and conserved 5'-haplotype, see Fig. 5.17.

The pGL3-Promoter vector (Promega) contains an SV40 promoter cloned upstream of the firefly luciferase gene (*Fluc*); DNA fragments suspected of containing putative enhancer elements can be examined by sub-cloning them upstream or downstream from the SV40 promoter and assaying for transcriptional activity. Using the synthetic oligonucleotides that were designed for EMSA analysis of the -3432poly-G polymorphism (see method section 3.2.18, Table 3.2) a series of pGL3-Promoter constructs were devised that contained two concatenated copies of either allele SNPG1 [pGL3-P(SNPG1)$_2$], allele SNPG2 [pGL3-P(SNPG2)$_2$], allele $(G)_{10}$ [pGL3-P(G10)$_2$], allele $(G)_{11}$ [pGL3-P(G11)$_2$], the consensus sequence for SP1[pGL3-P(SP1)$_2$] or an oligonucleotide containing a mutated SP1 binding site [pGL3-P(SP1Mut)$_2$] cloned upstream of the SV40 promoter (see method section 3.2.8.3 for details of construct formation; a schematic illustration of the constructs is shown in Fig. 5.17a). These constructs were co-transfected into OC-2 cells with either pSi-SP1 or pSi-SP3 expression vectors and activity was compared to basal levels in the presence of the empty pSi expression vector (Fig. 5.17b). The phRL-null vector (Promega) was used as an internal control to adjust for intra-experimental variation.

This line of analysis showed that in the absence of exogenous SP1 or SP3, basal activity from pGL3-Promoter constructs containing two concatenated copies of poly-G alleles SNPG1, SNPG2, $(G)_{10}$ or $(G)_{11}$ is very similar and does not differ appreciably from that of the empty pGL3-Promoter vector alone (Fig. 5.17b). Similarly, in the presence of pSi-SP3 expression vector activity of each poly-G- pGL3-promoter construct does not differ significantly compared to basal levels observed for each construct. In response to exogenous SP1 subtle increases in promoter activity are observed for constructs that contain high affinity SP1 binding alleles; pGL3-P(SNPG1)$_2$ and pGL3-P(G10)$_2$ exhibit 34% and 21% more activity respectively, compared to the empty pGL3-Promoter vector alone. In contrast, constructs pGL3-P(SNPG2)$_2$ and pGL3-P(G11)$_2$ that contain much weaker affinity SP1 binding alleles show negligible increases in activity; only 6% in each case. In comparison, in the presence of pSi-SP1 expression vector activity of constructs pGL3-P(SP1)$_2$ and pGL3-P(SP1Mut)$_2$ that contain two concatenated copies of the SP1 consensus sequence and a mutated SP1 binding site respectively, does not differ significantly from each other nor from that of the empty pGL3-Promoter vector alone. It is possible that this lack of enhanced activity from construct pGL3-P(SP1)$_2$ containing the SP1 consensus sequence may in part, be due to the fact that the OC-2 cell line expresses SP1 under standard culture conditions (see Fig. 5.12, page 201) and hence, masks any effect of adding exogenous SP1 (pSi-

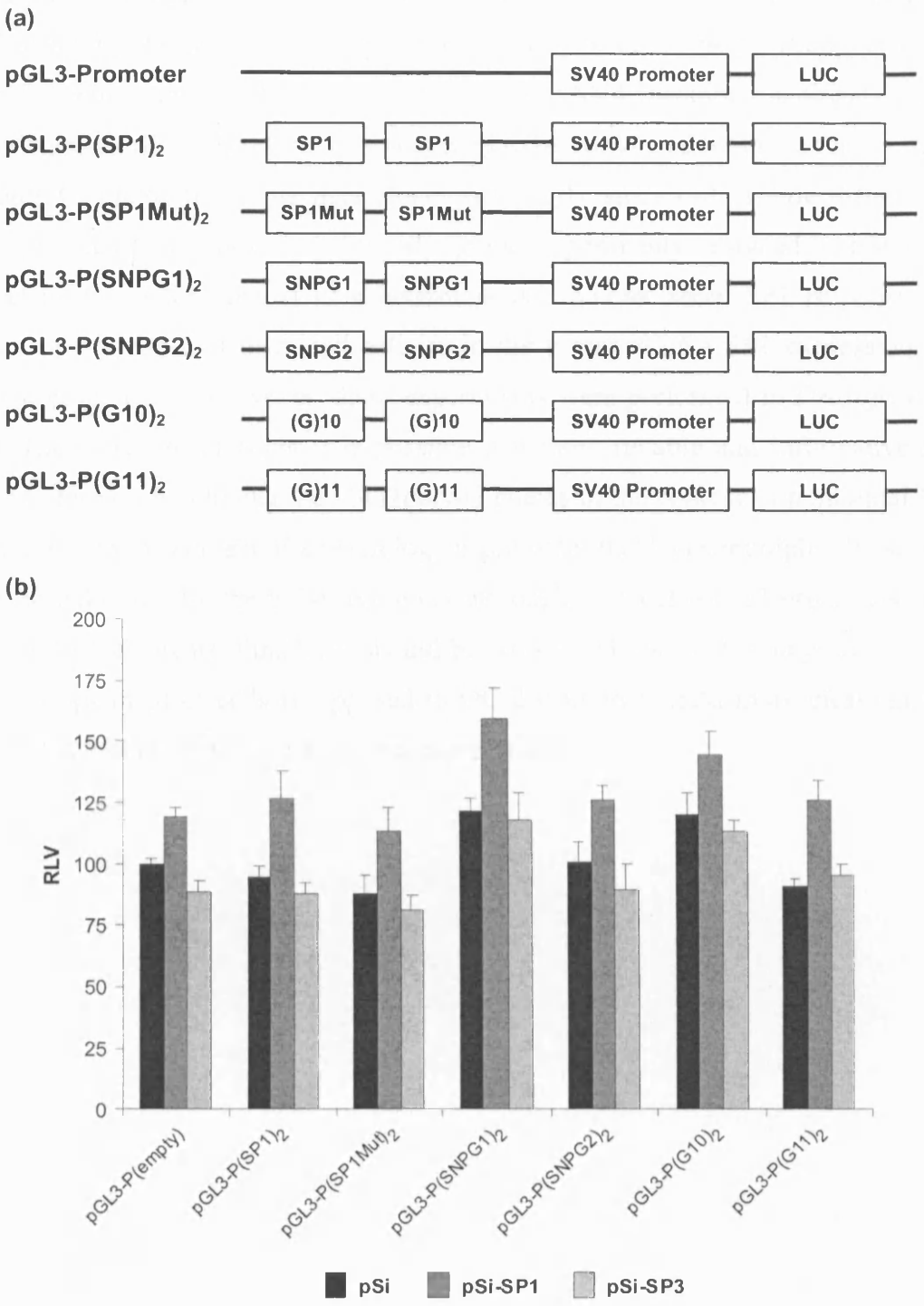


Figure 5.17 Effect of the -3432 poly-G polymorphism in the context of a heterologous promoter. (a) Schematic diagram (not to scale) to illustrate pGL3-Promoter luciferase reporter gene constructs that differ in the nature of the -3432 poly-G allele. Constructs contain two concatenated copies of either the SP1 consensus sequence, a mutated SP1 binding site, allele SNPG1, SNPG2, $(G)_{10}$ or $(G)_{11}$ cloned upstream of the heterologous SV40 promoter in pGL3-promoter vector (Promega). **(b)** OC-2 cells were co-transfected with 500ng of either pGL3-P(empty), pGL3-P$(SP1)_2$, pGL3-P$(SP1Mut)_2$ pGL3-P$(SNPG1)_2$ pGL3-P$(SNPG2)_2$ pGL3-P$(G10)_2$ or pGL3-P$(G11)_2$ in the presence of 500ng of either SP1 expression vector (pSi-SP1), SP3 expression vector (pSi-SP3) or the empty pSi expression vector. Renilla expression vector phRL-null (10ng) was used as an internal control to normalise for differences in transfection efficency. Error bars on charts represent the standard error of the mean. Luciferase activities are expressed relative to pGL3-P(empty) in the absence of stimulation by assigning a relative value of 100 (RLV). The experiment was performed in triplicate and repeated at least twice with two different preparations of DNA.

SP1 expression vector) to these cells. Moreover, it is possible that these results are confounded by the fact the SV40 promoter present in vector pGL3-Promoter contains several binding sites for SP1 thus, meaning that the SV40 promoter is already activated by endogenous SP1. Therefore, it is very hard to be convinced of any specific effects of the poly-G alleles from this data alone particularly since only subtle differences in promoter activity are observed. Indeed, similar experiments designed to test putative SP1 elements in the context of a heterologous promoter show SP1 responsive sites exhibit up to nine-fold increased activity in the presence of a SP1 expression vector (Hirano et al, 1998). However, these experiments were performed in *Drosophila* SL-2 cells. Therefore, in hindsight it is possible that more reliable and informative results may be obtained by cloning the -3432poly-G alleles in the context of a minimal Brn-3c promoter or in the context of a heterologous promoter that lacks multiple SP1 sites such as those exhibited by the SV40 promoter although, most if not all promoters contain putative SP1 elements. Finally, it should be considered that performing these transient transfections in SL-2 cells as opposed to OC-2 cells may yield more clear-cut results due to a lack of endogenous SP protein expression.

5.8 General discussion.

5.8.1 Limitations of functional data generated *in-vitro*.

The results of the EMSA analysis and transient transfection assays suggest that the -3432poly-G polymorphism in the Brn-3c promoter is a functional polymorphism. In particular protein-DNA investigations show that the nature of the poly-G allele modulates the binding of the SP1 transcription factor and there is convincing evidence that this alters the activity of the Brn-3c promoter when SP1 levels are limiting. However, it is important to remember that these results relate to an *in-vitro* context. Specific binding of a trans-acting factor to a cis-acting element *in-vitro* even when supported by functional studies to show that the cis-acting element in question is transcriptionally active, does not mean that the same applies *in-vivo*. A binding site *in-vitro* may not be accessible *in-vivo*. EMSA studies are performed with DNA fragments of limited length these same regions *in-vivo* may be masked by the presence of histones and other DNA-binding proteins and accordingly may not be available for binding. To assess whether a specific transcription factor binds to a specific DNA sequence *in-vivo* chromatin immunoprecipitation (ChIP) assays can be performed using a primary antibody specific to the transcription factor in question (see Shang et al, 2000; Iype et al, 2005; for reviews see Spencer et al, 2003 and Das et al, 2004). This line of analysis would be useful to yield information on the chromatin context of the -3432poly-G polymorphic site in the Brn-3c promoter *in-vivo*; to determine whether the -3432poly-G polymorphic site is accessible for transcription factor binding *in-vivo* and occupied by the endogenous SP1 protein as the EMSA analysis suggest (section 5.3.2). However, this line of analysis is severely hindered by the lack of human sensory hair cell lines and the difficulties in obtaining primary hair cells from the human inner ear.

In a similar manner, transient transfections to investigate promoter activity are performed with a DNA fragment of a defined length. These are isolated from their normal chromatin context and the influence of distal enhancer or inhibitory elements that could further modulate promoter activity are excluded from analysis. Indeed, enhancer activity has been reported to influence transcription at a distance as great as 85Kb from the transcription start site (Blackwood et al, 1998). Therefore, it should be considered that the size of the Brn-3c promoter fragments used in the luciferase reporter gene assays to examine the effect of the -3432poly-G allele on Brn-3c promoter activity is 3.6Kb and this may have limited the magnitude of the response observed; all promoter elements needed for expression of Brn-3c may not be present.

Chromatin structure is central to the regulation of gene expression (for reviews see Kingston et al, 1996; Lee and Young et al, 2000, see also Biggin, 2001). The basic structural monomer of chromatin is the nucleosome comprising a core of 146bp of DNA wrapped around an octamer of histone molecules; it is the packaging of DNA into nucleosomes that can repress or potentiate gene expression depending on the promoter context (for review see Lee and Young, 2000). Hence, in this regard it should also be considered that transient transfection assays are limited in that the exogenous promoter-reporter gene constructs in the form of plasmid DNA are not integrated into the cellular genome of the host cell but instead are maintained extra-chromosomally. Some shortcomings typical of transient transfection assay can be rectified by generation of stable transfectants although this is a very time consuming approach (for review see Knight, 2003). Stable transfectants are integrated into the host genome, are "subject to position effects from the surrounding chromatin" and accordingly are more typical of the *in-vivo* DNA packaging than their transient counterparts (Smith and Hager, 1997). However, plasmid integration into the genome of a host cell is a random event and not necessarily representative of the native chromatin site *in-vivo*. Reporter gene expression can be subject to anomalous regulation by the influence of host genomic regulatory elements near the site of plasmid integration that are absent from the native chromatin site. Consequently, this approach requires analysis of many clones in order to determine real effects and the value of this approach when looking for subtle differences in expression between two alleles is limited as variation between clones may mask such subtle effects (for review see, Knight, 2003). Analysis of mRNA expression profiles in different individuals that each carry a distinct genotype for the -3432poly-G polymorphism would allow one to conclude with more certainty whether differences in -3432 genotype manifest in differential expression of Brn-3c *in-vivo*. However, it is not possible to perform this line of analysis, as hair cells are experimentally inaccessible.

Thus, although transient transfection assay in OC-2 cells to characterise common sequence variants in the Brn-3c promoter is not a substitution for *in-vivo* functional analysis it is clear that this line of analysis can undoubtedly give valuable insight into the functional effect of sequence variation on Brn-3c regulation.

5.8.2 Limitations of analysis using MatInspector software: identification of nuclear protein C as an example.

Analysis using MatInspector software to identify putative transcription factor binding sites within a particular DNA sequence is limited by the comprehensiveness of the database at the time the analysis is performed. Transcription factors yet to be identified and transcription factors identified but lacking experimental verification of consensus binding site sequence are both unaccounted for. With this in mind the analysis using MatInspector software was repeated on the poly-G allele sequences during the preparation of this thesis to see if any recent additions to the database would yield more informative results as to the identity of nuclear protein C (see section 5.3.3). Results from this more recent search revealed that the transcription factor ZIC2 exhibits a high matrix similarity to all poly-G allele sequences examined except alleles SNPG1 and SNPG5 (see Appendix C). These results are of particular interest as the results of the zinc titration experiment (Fig. 5.10, page 195) shows that both poly-G alleles SNPG1 and SNPG5 appear to have a greater binding affinity for nuclear protein A compared to nuclear protein C. In contrast the remaining poly-G alleles appear to have comparable affinities for nuclear proteins C and A, at least in the presence of zinc.

ZIC2 belongs to the ZIC multigene family of zinc finger transcription factors that play a crucial role in neural development (for review see Aruga, 2004; Grinberg and Millen, 2005). Expression studies on the chick inner ear suggest that the ZIC gene family plays an important role in inner ear development (Warner at al, 2003). Interestingly, in the chick basilar papilla differential display analysis shows that ZIC2 expression is up-regulated after exposure to acoustic trauma (Gong et al, 1996). Genes that are differentially regulated after noise exposure in the chick inner ear may be involved in hair cell repair and survival. Evidence suggests that Brn-3c functions as a pro-survival factor for hair cells of the inner ear (Erkman et al, 1996; Xiang et al, 1997 and Xiang et al, 1998). Therefore, if nuclear protein C is in fact, ZIC2 it would be interesting to investigate whether ZIC2 is involved in Brn-3c regulation in hair cells and whether there is a connection between ZIC2 response to acoustic trauma and the pro-survival role of Brn-3c. Furthermore, the possibility that OC-2 derived nuclear protein C is ZIC2 is of particular interest when coupled with the fact that ZIC2 has been shown to competitively interact with SP1 and SP3 for binding site occupation within the 5'-flanking region of the human D_{1A} dopamine receptor gene (Yang et al, 2000). The -3432poly-G polymorphism constitutes a SP1 and SP3 binding site and the nature of the poly-G allele is a determinant of SP1 and SP3 binding (see Fig. 5.7 page 187; Fig 5.8,

page 189 and Fig. 5.15, page 207). Therefore, it would be interesting to investigate whether ZIC2 may also interact at this site in the Brn-3c promoter although it was not possible to perform this analysis within the time constraints of this project.

5.9 Conclusion

In conclusion, the -3432poly-G polymorphism in the Brn-3c promoter modifies the binding affinity of at least two OC-2 derived nuclear proteins, A and C. EMSA analysis is highly indicative that nuclear protein A is the transcription factor SP1 but attempts to identify nuclear protein C were not as successful. However, EMSA analysis suggests that nuclear protein C is a zinc finger transcription factor and it is clear that nuclear protein C is distinct to nuclear protein A. Further investigation incorporating purified human recombinant SP1 protein (Promega) in EMSA analysis and Brn-3c promoter-luciferase reporter gene constructs in transient transfection assay in OC-2 cells suggests that the -3432poly-G polymorphism in the Brn-3c promoter is a functional polymorphism; the nature of the poly-G allele modulates the binding of SP1 and there is convincing evidence that this alters the regulation of the Brn-3c promoter when SP1 levels are limiting, $p<0.05$ (paired t-test). It is well established that SP1 can competitively interact with SP3 for binding to the same binding sites (for reviews see Suske, 1999 and Kaczynski et al, 2003) and that for some SP1-responsive genes the nuclear SP1:SP3 ratio is an important determinant of gene regulation (Hoppe and Francone, 1998; Hata et al, 1998). With this in mind, attempts were made to establish whether SP1 and SP3 competitively interact at the -3432 polymorphic site in the Brn-3c promoter but it was not possible to ascertain whether this is the case. However, from EMSA analysis using *in-vitro* translated SP3 it is clear that the nature of the poly-G allele is a determinant of SP3 binding and transient transfection assays incorporating an SP3 expression vector suggest this has an effect, albeit subtle, on Brn-3c promoter activity although not at significantly different levels when tested by paired t-test analysis. Collectively, these findings raise the possibility that the -3432poly-G polymorphism in the Brn-3c promoter may be a risk factor for late onset hearing loss exhibited by a large proportion of the ageing population; this line of analysis is discussed further in Chapter 8 as part of a preliminary association study. The physiological significance of SP1 on Brn-3c regulation in hair cells is not known at present. The role of SP1 in Brn-3c regulation is explored further in the next Chapter by analysis incorporating the effect of the -566$(GT)_n$ repeat polymorphism on Brn-3c promoter activity.

6.0 Functional Characterisation Of Dinucleotide Repeat Sequence Variation At The Brn-3c Locus.

6.1 Introduction.

DNA sequence variation can manifest in many forms from single nucleotide substitutions to DNA elements that vary only by changes in the number of sequence repeats. Sometimes the variation is much more complex with DNA elements comprising both of these features such as the -3432poly-G polymorphism identified in the Brn-3c promoter (see Chapter 5). Repetitive DNA elements, also known as simple sequence repeats (SSRs) are tandem repeats of nucleotide motifs. Those that arise from short motifs; interactions of 1-6bp the mono-, di-, tri-, tetra- and penta-nucleotide repeats are commonly known as microsatellites. Whereas, those that form longer motifs, up to 60bp are termed minisatellites. The former are generally less than 100bp long; repeat number ranging from 2 up to a couple of dozen. Whereas, the later can reach up to 20Kb in length with repeat number extending into the hundreds (for reviews see Kashi et al, 1997; Li et al, 2002; Ellegren, 2004). The majority of microsatellite sequence variation arises with the short dinucleotide repeating motifs and is generally multi-allelic with a bimodal allelic distribution. The most abundant dinucleotide repeat is the $(CA / TG)_n$ motif; estimated to occur every 20-30Kb in mammals (Stallings et al, 1991). Dinucleotide microsatellites as with the mono-, tri-, tetra- and penta-microsatellites are mainly located within non-coding regions of the eukaryotic genome and by comparison are relatively rare within regions that encode proteins (for review see Li et al, 2002). Trinucleotide repeats are the exception to this rule; they are often found within protein-coding regions and for some time have known to be linked with disease (Stallings, 1994). The neurodegenerative disorder Huntington's disease is a good example; caused by an expansion of the triple repeat (CAG) in the coding region of the Huntington's disease gene that is translated into expanded tracts of polyglutamine (for review see Zoghbi and Orr, 2000).

Originally microsatellites were considered to be just 'junk' DNA or evolutionarily neutral genetic markers but there is now increasing evidence that some microsatellites play a functional role. Their involvement has been documented in a range of cellular processes to include gene regulation, chromatin organisation and

regulation of DNA metabolic processes such as recombination (for review see Li et al, 2002). Indeed, there is substantial evidence that polymorphic dinucleotide repeats located in non-coding regions have quantitative effects on gene expression able to both repress (Wu et al, 1994 and Gebhardt et al, 1999) and enhance gene expression (Akai et al, 1999, Borrmann et al, 2003 and Tadokoro et al, 2004). However, the exact mechanism as to how they do so is still not well understood. The ability to adopt alternative secondary structure conformations such as Z-DNA (a left-handed double helix which lacks the major groove characteristic of the common right-handed B-DNA double helix typically exhibited by the DNA sequence) appears to play an important role (Rothenburg et al, 2001). Alternating $(CG)_n$ repeats most readily form Z-DNA, followed by $(CA / TG)_n$ and $(TA)_n$ repeats (Rich et al, 1984). In addition, evidence suggests that certain microsatellites may act as cis-regulatory elements by the binding of transcription factors (Albanese et al, 2001; Contente et al, 2002; Iglesias et al, 2004). In many of these cases repeat number plays a key role; the ability of p53 to strongly transactivate the p53-induced gene 3 (PIG3) is dependent on the length of a pentanucleotide repeat in the PIG3 promoter to which p53 binds (Contente et al, 2002). Furthermore, there is accumulating evidence that polymorphic dinucleotide repeats that modulate gene regulation may contribute to inter-individual differences in susceptibility to disease (Searle et al, 1999, Yamada et al, 2000, Kodama et al, 2004). For example, a polymorphic dinucleotide $(GT)_n$ repeat in the 5'-flanking region of the human heme oxygenase-1 (HO-1) gene that has a functional affect on HO-1 gene transcription is associated with chronic pulmonary emphysema susceptibility (Yamada et al, 2000).

Three polymorphic dinucleotide repeats have been identified in the Brn-3c 5'-flanking region which appear common in the general population: two $(GT)_n$ dinucleotide repeats at positions -566 and -3495 and a $(GA)_n$ dinucleotide repeat at position -3457 (see section 4.5 and Fig. 4.16, page 162). It is possible that one or a combination of these dinucleotide repeats could have quantitative effects on Brn-3c expression. Indeed, the human type I collagen α2 (COL1A2) gene is regulated by the combination of two polymorphic dinucleotide repeats: one in the 5'-flanking region and one in the first intron of the COL1A2 gene (Akai et al, 1999). Therefore, in the search to identify common sequence variants at the Brn-3c locus that are functional and which may be a risk factor for late onset hearing loss exhibited by a large proportion of the ageing population the three polymorphic dinucleotide repeats at position -566, -3457 and -3495 in the Brn-3c 5'-flanking region were considered good candidates for functional analysis.

6.2 Strategy to investigate the significance of dinucleotide repeat sequence variation at position -566, -3457 and -3495 on Brn-3c regulation.

There is increasing evidence that polymorphic dinucleotide repeats located in non-coding regions can modulate gene regulation as discussed (section 6.1). Three polymorphic dinucleotide repeats have been identified in the Brn-3c 5'-flanking region: -566$(GT)_{17\text{-}23}$, -3457$(GA)_{1\text{-}3}$ and -3495$(GT)_{15\text{-}21,\ 24}$ raising the possibility that either or a combination of these repeats may have quantitative effects on Brn-3c expression. The -3457$(GA)_n$ repeat and the -3495$(GT)_n$ repeat are located immediately adjacent to each other in the Brn-3c 5'-flanking region; the -3457$(GA)_n$ repeat begins immediately at the 3-prime end of the -3495$(GT)_n$ repeat (see section 4.5, Fig. 4.16, page 162). From the initial PCR-SSCP screen and subsequent genotyping analysis of the -3432poly-G polymorphism for a preliminary association study it was clear that the -3495$(GT)_n$ repeat and -3457$(GA)_n$ repeat appeared in strong linkage disequilibrium with the -3432poly-G polymorphism thus forming native haplotypes 5' of the -3432 poly-G polymorphism (see section 4.4.2, Table 4.4, page 160 and section 4.5, Fig. 4.16, page 162). Due to the complexity of this region of the Brn-3c promoter upstream of the -3432poly-G polymorphism resulting from the high amount, the proximity and the repetitive nature of the variation identified complete characterisation of haplotypes at positions -3432, -3457 and -3495 in all individuals examined was not possible (this is discussed further in section 4.4.2, page 158). However, where it was possible to decipher haplotypes were of definite combinations as far as could be determined from the analysis performed. For example, if poly-G allele SNPG1 was present at position -3432 then the corresponding native 5' haplotype consisted of -3457$(GA)_2$ and -3495$(GT)_{15}$. An altered haplotype configuration was not observed 5' of the poly-G polymorphism if allele SNPG1 was present. Similarly, if poly-G allele SNPG2 was present at position -3432 the native 5' haplotype consisted of -3457$(GA)_2$ and -3495$(GT)_{20\ or\ 21}$. The -3457$(GA)_3$ was very rare. To emphasise, a long -3495 GT repeat length was never observed in conjunction with allele SNPG1 and conversely a short -3495 GT repeat length was never observed in concert with allele SNPG2. Haplotypes that it was possible to identify are summarised in Table 4.4, section 4.4.2, page 160.

Given the repetitive nature of the sequence variants at -3432, -3457 and -3495 and their proximity any attempts to alter appropriate Brn-3c promoter-luciferase

reporter gene constructs by site directed mutagenesis would prove extremely problematic if not impossible. Furthermore, no suitable restriction enzyme sites are found between these sequence variants to facilitate sub-cloning of alternative haplotype combinations by restriction digest. The complexity of this region of the Brn-3c promoter in the vicinity of the -3432poly-G polymorphism limited the nature of functional analysis that could be performed (a similar discussion of the difficulties encountered with functional analysis is presented in section 5.4, page 197). Therefore, with these caveats in mind it was decided to assess the significance of the -566, -3457 and -3495 dinucleotide repeats on Brn-3c promoter activity by examining the function of the dinucleotide $(GT)_n$ repeat at position -566 in the context of the -3432poly-G polymorphism and native 5' haplotype at -3457 and -3495.

The results of the PCR-SSCP screen of the Brn-3c locus on 45 individuals suggests that the -566 GT dinucleotide repeat in the Brn-3c promoter is a common sequence variant and repeat alleles ranging in length from 16 to 23 repeats were identified through subsequent sequencing analysis (see section 4.4.1, page 152). However, the -566$(GT)_n$ repeat was not genotyped in our sample cohorts as part of the preliminary association study as genotyping the -3432poly-G polymorphism constituted a large part of this project (see Chapter 8). Therefore, it was not known which of the alleles (n = 16 to 23) identified from preliminary analysis of the -566$(GT)_n$ repeat were the most common in the general population. Moreover, with the exception of the original PAC clone from which a 3.6Kb fragment of the Brn-3c promoter was isolated, it was not known in which context the -566$(GT)_n$ repeat existed in nature with the -3432poly-G polymorphism and native 5' haplotype at -3457 and -3495. Furthermore, of these combined haplotypes incorporating the -566$(GT)_n$ repeat, -3432poly-G polymorphism, -3457$(GA)_n$ repeat and -3495$(GT)_n$ repeat it was not known which were the most common in the general population. The complexity of the Brn-3c promoter 5' of the -3432poly-G polymorphism made identification of the *naturally* occurring *common* haplotypes at -3495, -3457, -3432 and -566 technically difficult. Therefore, it was decided to devise controlled haplotypes 5'of the -566$(GT)_n$ repeat based on the limited allele information known and to examine the significance of the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5' haplotype (at -3495 and -3457) under these conditions.

In some cases it is clear that the actual dinucleotide repeat number is a determinant of the effect of the dinucleotide repeat on gene regulation (Gebhardt et al, 1999; Shimajiri et al, 1999). Often, the two extremes in dinucleotide repeat length have

pronounced effects. For example, using a nuclear run-off assay, transcriptional activity of the epidermal growth factor receptor (EGFR) gene was found to decline with increasing length of a highly polymorphic dinucleotide $(CA)_n$ repeat in the first intron of this gene; in particular the longest allele examined (n = 21) inhibited transcription of the EGFR gene by 80% compared to the shortest allele examined (n = 16) (Gebhardt et al, 1999). Similarly, luciferase-reporter gene assays designed to assess the importance of a dinucleotide $(CA)_n$ repeat at position -90 in the promoter of the matrix metalloproteinase 9 (MMP-9) gene showed that in oesophageal carcinoma cell lines an MM-9-luciferase reporter gene construct that carried the $(CA)_{14}$ allele exhibited only 50% of the activity of the MM-9-luciferase reporter gene construct that carried the $(CA)_{21}$ allele (Shimajiri et al, 1999). Consequently, when designing Brn-3c promoter-luciferase reporter gene constructs to examine the significance of the -566 dinucleotide $(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5' haplotype, emphasis was placed on the two extremes of the -566$(GT)_n$ repeat length: 16 and 23 repeats.

The common -3432poly-G alleles SNPG1 and SNPG2 that were originally identified as part of the PCR-SSCP screen have native 5' haplotypes that consist of a short and a long -3495$(GT)_n$ repeat, respectively (see Table 4.4, section 4.4.2, page 160). Therefore, poly-G alleles SNPG1 and SNPG2 and their native 5' haplotypes at -3457 and -3495 were selected to form the distal 5' haplotypes of Brn-3c promoter-luciferase reporter gene constructs designed to assess the significance of the -566 dinucleotide $(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5' haplotype. Using luciferase reporter vector pGL3-Basic (Promega) and a 3.6Kb Brn-3c promoter fragment that contained allele SNPG1 and native 5' haplotype of -3457$(GA)_2$: -3495$(GT)_{15}$ a series of constructs were generated on a common backbone to carry different lengths of the dinucleotide $(GT)_n$ repeat at position -566 (see method section 3.2.8.1 for details of construct formation). Specifically, repeat number varied from 23 repeats (p-Luc-SNPG1$_{-566GT23}$) to 20 repeats (p-Luc-SNPG1$_{-566GT20}$), to 16 repeats (p-Luc-SNPG1$_{-566GT16}$) to represent constructs containing the longest, intermediate and shortest lengths of the -566 dinucleotide $(GT)_n$ repeat, respectively (a clear illustration of the variation present in the constructs is given Fig 6.1a, page 228; of these constructs, construct p-Luc-SNPG1$_{-566GT20}$ contains a naturally occurring haplotype at -3495, -3457, -3432 and -566 as it was derived from the original PAC clone containing a 3.6Kb fragment of the Brn-3c promoter). Similarly using poly-G allele SNPG2 with native 5' haplotype -3457$(GA)_2$: -3495$(GT)_{21}$, the dinucleotide $(GT)_n$ repeat at position -566 was varied from 23 (p-Luc-SNPG2$_{-566GT23}$) to 16 repeats (p-Luc-SNPG2$_{-566GT16}$) (a clear

illustration of the variation present in the constructs is given see Fig. 6.1a, page 228; see method section 3.2.8.1 for details of construct formation). This formed five different 3.6Kb Brn-3c promoter-luciferase reporter gene constructs that each carried a distinct 5' haplotype of the Brn-3c promoter for use in functional analysis.

6.3 The significance of -566(GT)$_n$ repeat sequence variation on Brn-3c promoter activity in the context of the -3432poly-G polymorphism and native 5' haplotype.

In an attempt to examine the significance of the -566(GT)$_n$ repeat on Brn-3c promoter activity in the context of the -3432poly-G polymorphism and native 5' haplotype five different luciferase reporter gene constructs that each carried a distinct 5' haplotype of the Brn-3c promoter were generated: p-Luc-SNPG1$_{-566GT23}$, p-Luc-SNPG1$_{-566GT20}$, p-Luc-SNPG1$_{-566GT16}$, p-Luc-SNPG2$_{-566GT23}$ and p-Luc-SNPG2$_{-566GT16}$ (see Fig. 6.1a). Notably in these constructs the length of the dinucleotide (GA)$_n$ repeat at position -3457 did not vary; both allele SNPG1 and SNPG2 have a native -3457(GA)$_n$ repeat number of two. Accordingly, the effect, if any, of the -3457 dinucleotide (GA)$_n$ repeat on Brn-3c promoter activity could not be assessed in these experiments.

6.3.1 Effect of -566(GT)$_n$ repeat sequence variation on basal activity of the Brn-3c promoter in the context of the -3432poly-G polymorphism and native 5' haplotype.

To assess whether the length of the -566(GT)$_n$ repeat allele has an effect on basal activity of the Brn-3c promoter each of the five Brn-3c promoter-luciferase reporter gene constructs carrying a unique 5' haplotype of the Brn-3c promoter were transfected into OC-2 cells along with the pRL-SV40 vector (Promega) as an internal control to adjust for intra- and inter-experimental variation (Fig. 6.1b). The results of this analysis showed that changing the length of the -566(GT) repeat from 23 to 16 repeats has negligible effect on basal activity of the Brn-3c promoter if poly-G allele SNPG2 and the long native GT repeat at -3495 form the distal 5' haplotype (Fig. 6.1b; mean RLV: 281 ± 8.87 versus 286.06 ± 8.29, respectively). In contrast, varying the length of the -566(GT)$_n$ repeat from 23 to 20 to 16 repeats has a subtle effect on basal activity of the Brn-3c promoter when the distal 5' haplotype consists of poly-G allele SNPG1 and the short native GT repeat at -3495 (Fig. 6.1b). Most evidently, basal activity from construct p-Luc-SNPG1$_{-566GT20}$ that carries a -566(GT)$_n$ allele of 20 repeats is 24% greater compared to that from construct p-Luc-SNPG1$_{-566GT23}$ that carries a -566(GT) repeat length of 23 repeats (Fig. 6.1b; mean RLV: 338.04 ± 11.86 versus 272.09 ±7.34, respectively) and this difference is statistically significant, $p<0.05$ (t-test).

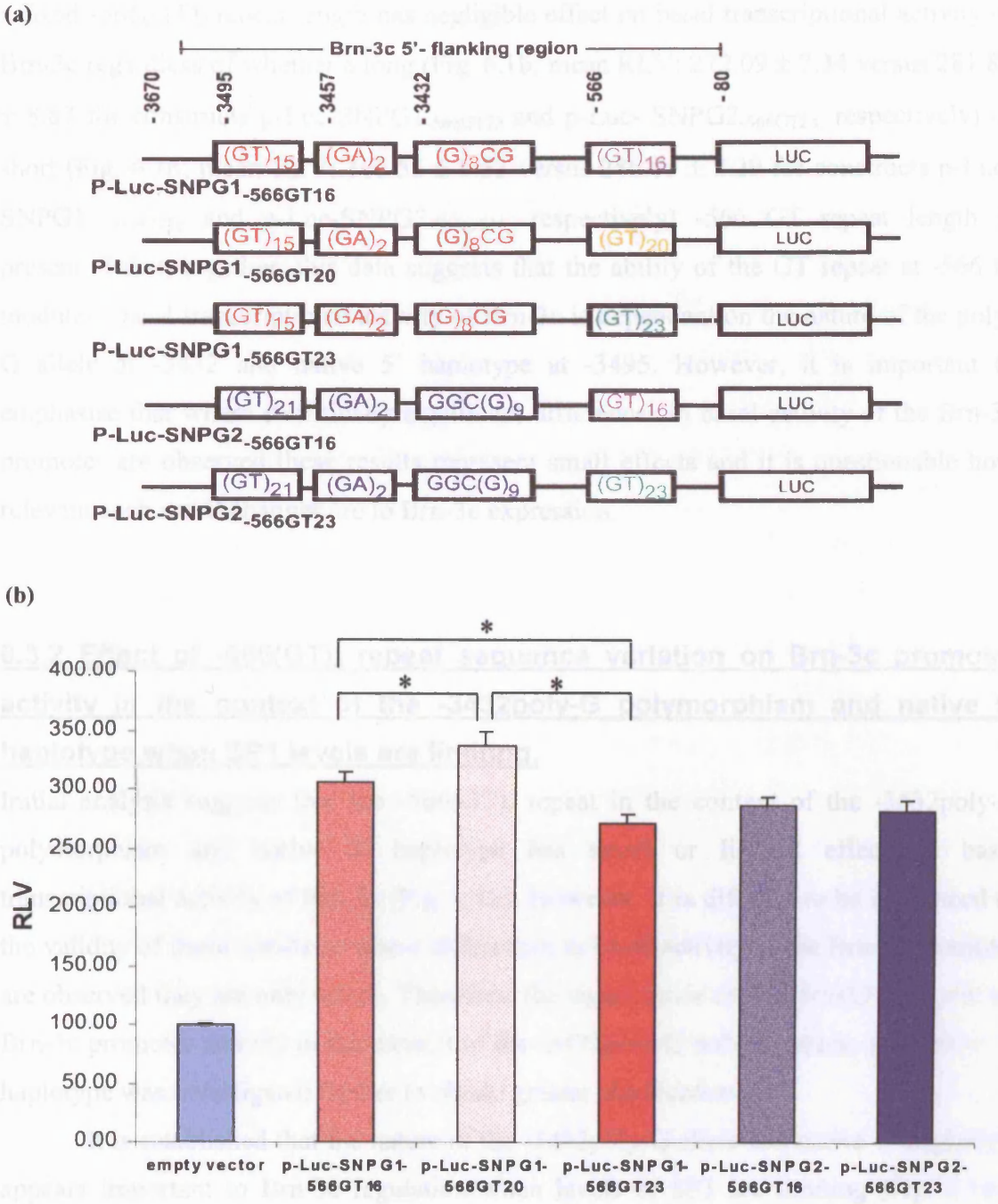


Figure 6.1 Effect of -566$(GT)_n$ repeat sequence variation on basal activity of the Brn-3c promoter in the context of the -3432poly-G polymorphism and native 5' haplotype. **(a)** A schematic diagram (not to scale) to illustrate the five 3.6Kb Brn-3c promoter-luciferase reporter gene constructs that each carry a distinct 5'haplotype of the Brn-3c promoter. -3432 poly-G allele SNPG1 and native 5' haplotype is highlighted in red whereas -3432poly-G allele SNPG2 and native 5' haplotype is highlighted in blue. The length of the -566 GT repeat was varied from 16 to 20 to 23 repeats as indicated. **(b)** OC-2 cells were transfected with 500ng of either p-Luc-SNPG1$_{566GT16}$, p-Luc-SNPG1$_{566GT20}$, p-Luc-SNPG1$_{566GT23}$, p-Luc-SNPG2$_{566GT16}$, p-Luc-SNPG2$_{566GT23}$ or the empty pGL3-Basic vector as indicated. The renilla expression vector pRL-SV40 (100ng) was used as an internal control to normalise for differences in transfection efficiency. Error bars on charts represent the standard error of the mean. Luciferase activities are expressed relative to the empty pGL3-Basic vector (RLV). The experiment was performed in triplicate and repeated at least twice with two different preparations of DNA. * $p<0.05$ (paired t-test; d.f. = 22). The data for constructs p-Luc-SNPG1$_{566GT23}$ and p-Luc-SNPG2$_{566GT23}$ is reproduced from Fig. 5.11b section 5.4, page 199.

It is also evident that varying the nature of the poly-G allele at -3432 in the presence of a fixed -566$(GT)_n$ repeat length has negligible effect on basal transcriptional activity of Brn-3c regardless of whether a long (Fig. 6.1b; mean RLV: 272.09 ± 7.34 versus 281.80 ± 8.87 for constructs p-Luc-SNPG1$_{-566GT23}$ and p-Luc- SNPG2$_{-566GT23,}$ respectively) or short (Fig. 6.1b; mean RLV: 305.55 ± 9.53 versus 286.06 ± 8.29 for constructs p-Luc-SNPG1$_{-566GT16}$ and p-Luc-SNPG2$_{-566GT16,}$ respectively) -566 GT repeat length is present. Taken together, this data suggests that the ability of the GT repeat at -566 to modulate basal transcriptional activity of Brn-3c is dependant on the nature of the poly-G allele at -3432 and native 5' haplotype at -3495. However, it is important to emphasise that where statistically significant differences in basal activity of the Brn-3c promoter are observed these results represent small effects and it is questionable how relevant such small changes are to Brn-3c expression.

6.3.2 Effect of -566$(GT)_n$ repeat sequence variation on Brn-3c promoter activity in the context of the -3432poly-G polymorphism and native 5' haplotype when SP1 levels are limiting.

Initial analysis suggests that the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5' haplotype has small or limited effect on basal transcriptional activity of Brn-3c (Fig. 6.1b). However, it is difficult to be convinced of the validity of these results as where differences in basal activity of the Brn-3c promoter are observed they are only subtle. Therefore, the significance of the -566$(GT)_n$ repeat on Brn-3c promoter activity in the context of the -3432poly-G polymorphism and native 5' haplotype was investigated further to obtain greater clarification.

It is established that the nature of the -3432poly-G allele and native 5' haplotype appears important to Brn-3c regulation when levels of SP1 are limiting (Fig. 5.11b, section 5.4.1, page 199). Transient transfection assays in OC-2 cells show that Brn-3c promoter-luciferase reporter gene constructs that carry allele SNPG1 (and native 5' haplotype) are significantly less sensitive to the effect of dominant negative SP1 mutant compared to constructs that carry allele SNPG2 (and native 5' haplotyep) $p<0.05$ (paired t-test; Fig. 5.13b, section 5.4.1, page 203). Moreover, EMSA analysis is highly suggestive that it is the nature of the poly-G allele at position -3432 in the Brn-3c promoter that co-ordinates this response (see section 5.3). Therefore, in an attempt to validate the subtle differences in basal activity of the Brn-3c promoter as observed when the -566$(GT)_n$ repeat is examined in the context of the -3432poly-G polymorphism and

native 5' haplotype (Fig. 6.1b) this line of analysis was extended to incorporate the effect, if any, of -566$(GT)_n$ repeat sequence variation on Brn-3c promoter activity in the context of the -3432poly-G polymorphism and native 5' haplotype under conditions where SP1 is limiting.

Brn-3c promoter-luciferase reporter gene constructs that each carry a distinct 5' haplotype of the Brn-3c promoter were transfected into OC-2 cells alongside increasing amounts of dominant negative SP1 expression vector, pSi-DBD-SP1 (see Fig. 6.2). The phRL-null vector (Promega) was used as an internal control to adjust for differences in intra- and inter-experimental variation. This line of analysis showed that Brn-3c promoter-luciferase reporter gene constructs that carry poly-G allele SNPG1 and native 5' haplotype are only sensitive to the effect of dominant negative SP1 mutant when the long form of the GT repeat is present at position -566. Increasing amounts of dominant negative SP1 expression vector decreases transcriptional activity of the Brn-3c promoter in a dose-dependent manner when -566$(GT)_n$ contains 23 repeats but no significant change in activity is observed if the length of the -566 GT repeat is modified to 16 or 20 repeats (Fig. 6.2b). Moreover, this effect is not limited to Brn-3c promoter-luciferase reporter gene constructs that carry allele SNPG1. Construct, p-Luc-SNPG2$_{-566GT23}$ that contains poly-G allele SNPG2 and native 5' haplotype together with the long form of the GT repeat at position -566 is extremely sensitive to the effect of dominant negative SP1 mutant, exhibiting 60% less activity in the presence of 4-fold excess dominant negative SP1 expression vector. Yet, the ability of dominant negative SP1 mutant to reduce expression in this manner is lost if the GT repeat length at position -566 is reduced from 23 to 16 repeats (Fig. 6.2b). These findings show that regardless of whether the poly-G allele at -3432 contains a high (allele SNPG1) or low (allele SNPG2) affinity SP1 binding site dominant negative SP1 mutant can only reduce transcriptional activity from Brn-3c promoter-luciferase reporter gene constructs if the long allele of the GT repeat is present at position -566. However, it is clear from the original analysis (see Fig 5.13b, section 5.4.1, page 203 and for comparison reproduced in Fig 6.2b) that in the presence of the long allele of the -566$(GT)_n$ repeat the actual extent of the reduction in Brn-3c promoter activity depends on the nature of the -3432 poly-G allele and native 5' haplotype. If the long allele of the $(GT)_n$ repeat is present at position -566 in concert with the low affinity SP1 binding allele, SNPG2, at -3432 Brn-3c promoter activity is reduced to a greater extent than if the 5' haplotype consists of the high affinity SP1 binding allele, SNPG1, at position -3432. Taken together, this data suggests that both the length of the GT repeat at position -566 and the nature of the

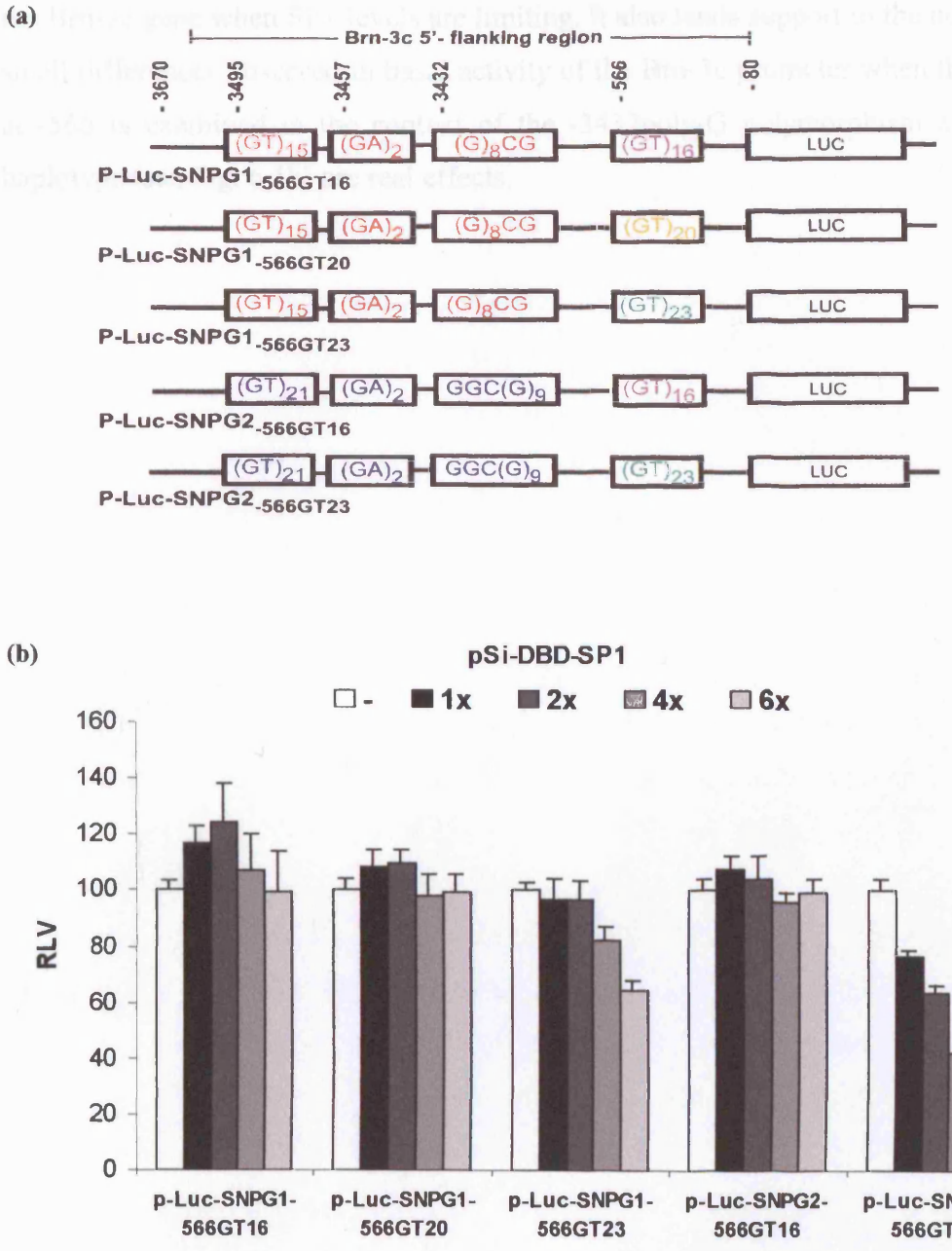


Figure 6.2 Effect of -566$(GT)_n$ repeat sequence variation on Brn-3c promoter activity in the context of the -3432poly-G polymorphism and native 5' haplotype when SP1 levels are limiting. **(a)** A schematic diagram (not to scale) to illustrate the five 3.6Kb Brn-3c promoter-luciferase reporter gene constructs that each carry a distinct 5'haplotype of the Brn-3c promoter. -3432 poly-G allele SNPG1 and native 5' haplotype is highlighted in red whereas -3432poly-G allele SNPG2 and native 5' haplotype is highlighted in blue. The length of the -566 GT repeat was varied from 16 to 20 to 23 repeats as indicated. **(b)** OC-2 cells were transfected with 500ng of either p-Luc-$SNPG1_{566GT16}$, p-Luc-$SNPG1_{566GT20}$, p-Luc-$SNPG1_{566GT23}$, p-Luc-$SNPG2_{566GT16}$, p-Luc-$SNPG2_{566GT23}$ in the presence of increasing amounts (0, 0.5, 1, 2 & 3μg) of dominant negative SP1 expression vector (pSi-DBD-SP1). The empty pSi expression vector was used to standardise DNA concentrations between wells and the renilla expression vector phRL-null (10ng) was used as an internal control to normalise for differences in transfection efficiency. Error bars on charts represent the standard error of the mean. Luciferase activities are expressed relative to each construct in the absence of stimulation by assigning a relative value of 100 (RLV). The experiment was performed in triplicate and repeated at least twice with two different preparations of DNA. The data for p-Luc-$SNPG1_{566GT23}$ and p-Luc-$SNPG2_{566GT23}$ is reproduced from Fig. 5.13b, page 203.

poly-G allele at -3432 and native 5' haplotype at -3495 are important in regulation of the Brn-3c gene when SP1 levels are limiting. It also lends support to the notion that the small differences observed in basal activity of the Brn-3c promoter when the GT repeat at -566 is examined in the context of the -3432poly-G polymorphism and native 5' haplotype (see Fig. 6.1b) are real effects.

6.4 General discussion.

From this study it is not possible to delineate the molecular mechanism(s) by which the -566$(GT)_n$ repeat coupled with the nature of the -3432poly-G allele and native 5' haplotype influences Brn-3c regulation, either under basal conditions (Fig. 6.1b) or when SP1 levels are limiting (Fig. 6.2b). Indeed, from the limited data presented in Fig. 6.1b it is not possible to draw a clear-cut relationship between Brn-3c 5' haplotype combination and corresponding basal activity of the Brn-3c promoter. Although, it is evident that constructs that carry poly-G allele SNPG1 at -3432 and native 5' haplotype are sensitive to a change in -566$(GT)_n$ repeat length compared to those that carry allele poly-G allele SNPG2 and native 5' haplotype (Fig. 6.1b). It should be considered that had a wider range of haplotype combinations been tested for constructs that carry poly-G allele SNPG2 at -3432 a difference in basal activity of the Brn-3c promoter may have been observed. However, it was not practical to test all possible haplotype combinations and selective haplotype combinations were chosen as discussed previously.

Interestingly, the basal activity of Brn-3c promoter-luciferase reporter gene constructs that carry poly-G allele SNPG1 at -3432 with the short native GT repeat at -3495 (p-Luc-SNPG1$_{-566GT16}$, p-Luc-SNPG1$_{-566GT20}$ and p-Luc-SNPG1$_{-566GT23}$) is not linear with respect to -566$(GT)_n$ repeat length (Fig. 6.1b). This suggests that it is the juxtaposition of poly-G allele SNPG1 at -3432 and native 5'haplotype with the -566$(GT)_n$ repeat allele that has an affect on basal activity of the Brn-3c promoter rather than the length of the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism. Indeed, similar data has been observed for the COL1A2 gene (Akai et al, 1999). Transcriptional activity of COL1A2 is enhanced by the presence of two dinucleotide repeats (one in the 5'-flanking region and one in the first intron of the COL1A2 gene) and different allelic combinations of repeat length modulate this effect, but the quantitative changes in COL1A2 promoter activity are not linearly proportional to the length of each dinucleotide repeat. Accordingly, it is proposed that the differences in COL1A2 promoter activity are due to the different allele combinations rather than the length of each repeat.

In comparison, the data presented in Fig. 6.2b is strongly indicative that the length of the -566$(GT)_n$ repeat in the context of the -3432 poly-G polymorphism and native 5' haplotype plays an important role in SP1-mediated regulation of Brn-3c. Specifically, when coupled with EMSA analysis (section 5.3) the data suggest that a functional interplay between the -566$(GT)_n$ repeat and the SP1 binding site in the -3432

poly-G sequence is important to Brn-3c regulation, at least under conditions where SP1 is limiting. From this study it is not possible to delineate the molecular mechanism by which this may occur however, this data is consistent with several putative models for which there are varying degrees of evidence in other genes and these will now be discussed.

6.4.1 Putative models to explain the functional interplay between the -566$(GT)_n$ repeat and the SP1 binding site in the -3432 poly-G sequence.

SP1 is an extremely well characterised transcription factor (for reviews see Phillipsen and Suske, 1999; Kaczynski et al, 2003; Chu and Ferro, 2005; Zhao and Meng, 2005). In addition to a well-known role in formation of the basal transcription complex (Gill et al, 1994), SP1 interacts with a multitude of other transcription factors (Perkins et al, 1994, Lin et al, 1996 and Lopez-Rodriguez et al, 1997) to play a more specific role in gene expression. It has been suggested that the ability of SP1 to confer flexibility to DNA upon binding SP1-recognition motifs, enabling the DNA to bend may facilitate its interaction with proteins that are spaced far apart on the DNA sequence (Sjottem et al, 1997). Therefore, it is tempting to speculate that the shorter allele of the -566$(GT)_n$ repeat could impair the binding of an, as yet unidentified, proximally bound transcription factor that associates with SP1 bound at the distal -3432 site and is crucial for SP1-mediated regulation of Brn-3c expression. It could be hypothesised that a proximal transcription factor may bind within the -566 GT dinucleotide repeat, itself. In agreement with this concept, the strength of nuclear protein binding to a polymorphic CA repeat within the promoter of the MMP-9 gene has been shown to be dependent on the length of the CA repeat; with longer CA repeat alleles showing a much higher binding affinity than their shorter counterparts (Shimajiri et al, 1999). Although, in this study the nuclear protein was not identified and it may not be a transcription factor. Had time permitted EMSA analysis could have been performed with different lengths of the -566$(GT)_n$ repeat to see if repeat length modified the binding of a transcription factor or factors. However, on the whole evidence that transcription factors can bind dinucleotide repeats is lacking, although they have been reported to bind tri-, tetra- and penta-nucleotide repeat motifs (Albanese et al, 2001; Contente et al, 2002; Iglesias et al, 2004). Alternatively, it could be hypothesised that a proximal transcription factor binds the DNA sequence adjacent to the -566 GT dinucleotide repeat. Certainly, it is feasible that an allele containing a shorter -566 GT repeat length could have a significant effect

on binding of an adjacent transcription factor as flanking DNA sequences can influence the context and configuration of cis-acting elements (for example see Dawson et al, 1996a).

On the other hand, it may not be the binding of a proximal transcription factor that is impaired by the shorter allele of the -566$(GT)_n$ repeat but rather just its association with SP1 bound at the distal -3432 site. This is possible, it has been suggested that dinucleotide repeats may serve to bring different regions of a promoter into close proximity (for review see Li et al, 2002). Hence, polymorphic dinucleotide repeats that vary in length may fold into different secondary conformations that do not readily favour protein-protein associations (for example see Gebhardt et al, 1999). DNA-bound SP1 is known to self-associate; looping out the intervening DNA such that SP1 bound at sites distal to the minimal promoter can exhibit synergistic interaction with proximally bound SP1 to enhance transcription (Su et al, 1991, Mastrangelo et al, 1991). Consequently, it is tempting to speculate that if SP1 bound at the -3432 site associates with a proximal transcription factor that this proximal transcription factor may be SP1 itself. In support of this concept recent experiments in our laboratory have identified a functional SP1 binding site within the minimal promoter region of Brn-3c (Jagutpal et al, unpublished data). This proximal SP1 binding site at position -345 is sufficiently close to the -566 GT dinucleotide repeat. Therefore, it is certainly possible that the shorter allele of the -566$(GT)_n$ repeat could hinder the binding or an association between SP1 at the proximal -345 site with SP1 bound at the distal -3432 site.

In summary, it is clear that the data suggest a functional interaction between the -566$(GT)_n$ repeat and the -3432poly-G polymorphism and native 5' haplotype which associate to determine the response of the Brn-3c gene to SP1 under conditions where SP1 is limiting. The molecular mechanism as to how this occurs is not known but it is clear that there are many possibilities, which are supported by evidence in other genes. The complexity and repetitive nature of the sequence variants identified (dinucleotide repeat sequence variation at -3495 and -566 coupled with the multi-allelic poly-G polymorphism at -3432) limited the nature of functional analysis that could be performed and hindered full investigation of these effects. However, collectively it is clear that there is evidence of common variation in the Brn-3c promoter and there is convincing evidence that this has a functional effect on regulation of the Brn-3c gene.

6.4 Conclusion.

In conclusion, transient transfection assay in OC-2 cells suggests that the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5'haplotype has an affect on basal transcriptional activity of Brn-3c. Although, where differences in basal activity of the Brn-3c promoter are observed between different luciferase-reporter gene constructs that each carry a distinct 5' haplotype of the Brn-3c promoter the differences are small and it is questionable how real these effects are. However, it is clear that the length of the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5'haplotype has an affect on Brn-3c regulation under conditions where SP1 is limiting. Moreover, when taken together with EMSA analysis (section 5.3.2, page 185) the data suggest that a functional interplay between the -566$(GT)_n$ repeat and the SP1 binding site in the -3432poly-G allele sequence is a determinant of this response. These results also lend support to the concept that subtle differences in basal activity of the Brn-3c promoter as observed when the -566$(GT)_n$ repeat is examined in the context of the -3432poly-G polymorphism and native 5'haplotype are real effects. Collectively, these results are suggestive that the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5' haplotype is functional.

The physiological significance of these results are unknown and require more investigation than was possible within the time constraints of this project to identify the signalling mechanisms that regulate SP1 activity in auditory hair cells. Furthermore, with the exception of construct p-Luc-SNPG1$_{-566GT20}$ which carries a naturally occurring haplotype spanning sequence variants: -3495$(GT)_n$, -3457$(GA)_n$, -3432poly-G polymorphism and -566$(GT)_n$ it is important to consider that these results are based on data generated using controlled haplotypes 5'of the -566$(GT)_n$ repeat. The complexity of the Brn-3c promoter 5' of the -3432poly-G polymorphism made identification of the *naturally* occurring *common* haplotypes at -3495, -3457, -3432 and -566 difficult and limited the nature of functional analysis that could be performed. It would be interesting to assess whether the common native haplotypes at these loci give similar results either under basal conditions or when SP1 levels are limiting although the genotyping required determining the naturally occurring common haplotypes would be technically challenging.

7.0 Functional Characterisation Of -1391 A>C At The Brn-3c Locus.

7.1 Introduction.

At the onset of this project an A >C single nucleotide substitution at position -1391 in the Brn-3c 5'-flanking region had been reported in the public NCBI SNP database, NCBI SNP Cluster ID: rs1368402 (see Table 4.1, section 4.1.1, page 131). However, this sequence variant had not been validated and the significance of this single nucleotide substitution with respect to Brn-3c regulation and / or expression was not known. Mutation scanning of the Brn-3c gene by PCR-SSCP analysis coupled with sequencing confirmed the presence of the -1391A >C substitution and suggested that the -1391 sequence variant is common in the general population with a rare allele frequency of 0.28 (based on SSCP analysis on 45 individuals; see Table 4.2, section 4.3.1, page 146). There is accumulating evidence that common sequence variants involving a single base change in the nucleotide sequence which manifest within gene regulatory regions underlie susceptibility to common, complex disease (Dawson et al, 1993; Zhang et al, 1999; Jormsjo et al, 2000; Niimi et al, 2002; see also section 1.6.1, page 49 and section 5.1, page 166). Consequently, the -1391A >C single nucleotide substitution in the Brn-3c 5'-flanking region could be a risk factor for late onset hearing loss exhibited by a large proportion of the ageing population; it may impair the binding of a transcription factor needed for regulation of the Brn-3c gene and thus lead to inter-individual variations in Brn-3c expression. Therefore, as part of the aim to asses whether common sequence variants in the Brn-3c gene are a risk factor for late onset hearing loss by first assessing whether they are functional, functional analysis was performed on the -1391 A >C substitution. This line of analysis is explored in this chapter.

7.2 Effect of the -1391 A >C single nucleotide substitution on binding of OC-2 nuclear proteins .

In order to assess whether the -1391A >C single nucleotide substitution modifies the binding of nuclear proteins to this region of the Brn-3c promoter EMSA analysis was performed using nuclear extracts from OC-2 cells. Synthetic oligonucleotides spanning each allele of the -1391 sequence variant in the Brn-3c promoter were devised and annealed to form double stranded probes: probes A and C representing wild type and variant alleles, respectively (see Table 3.2, method section 3.2.18).

In the first instance in order to assess whether the A to C substitution at -1391 in the Brn-3c promoter modifies nuclear protein binding probes A and C were incubated with 8-10μg of OC-2 nuclear extract in the absence of competitor DNA, Fig. 7.1. However, this approach did not reveal differential protein binding between alleles A and C (Fig. 7.1). Three main shifted protein-DNA complexes are observed at analogous positions on each probe, termed G, H and J (Fig. 7.1). The intensity of each shifted protein-DNA complex appears similar between each probe suggesting that alleles A and C have similar affinities for the protein or proteins involved in the formation of each complex. However, it is possible that a difference in affinity for the protein(s) involved

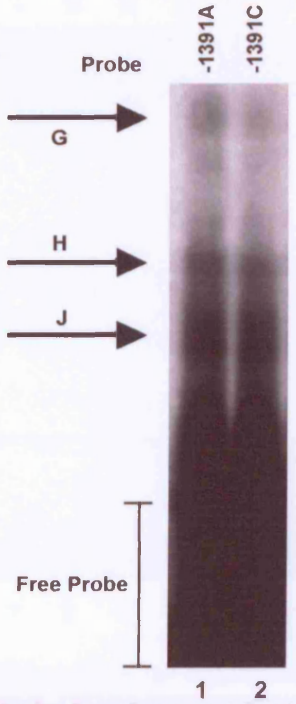


Figure 7.1 Binding of OC-2 derived nuclear proteins to alleles -1391A and -1391C. Radiolabelled oligonucleotides representing alleles -1391A (lane 1) and -1391C (lane 2) were incubated on ice with 8-10μg OC-2 nuclear extract. Three main shifted protein-DNA complexes are observed at analogous positions on each probe and are denoted by the arrows G, H and J. The intensity of each shifted protein-DNA complex appears similar between each probe suggesting that alleles A and C have similar affinities for the protein or proteins involved in the formation of each complex.

in the formation of complex G, H or J exists between alleles A and C but is too subtle to be observed directly in this manner (for simplicity one protein will be assumed to be involved in the formation of each complex hereafter referred to as protein G, H and J).

To test the possibility that a subtle difference in affinity for one of the proteins involved in the formation of complex G, H or J exists between alleles A and C increasing amounts (25-, 50-, 100-, 500- and 1000-fold excess) of unlabelled A probe (Fig. 7.2, lanes 3, 5, 7, 9 & 11) or unlabelled C probe (Fig. 7.2, lanes 4, 6, 8, 10 & 12) were used to compete with A probe for binding. This line of analysis showed that allele A is able to compete for nuclear protein J; the shifted complex J is reduced at 100-fold excess and is further reduced in intensity as the amount of unlabelled A probe is increased 500- fold (Fig. 7.2, comparing lane 1 with 7 and 9). Although, as evident in Fig. 7.2, lane 11 this competition does not proceed to completion; even at 1000-fold excess of the unlabelled A probe some nuclear protein J binding is still apparent on

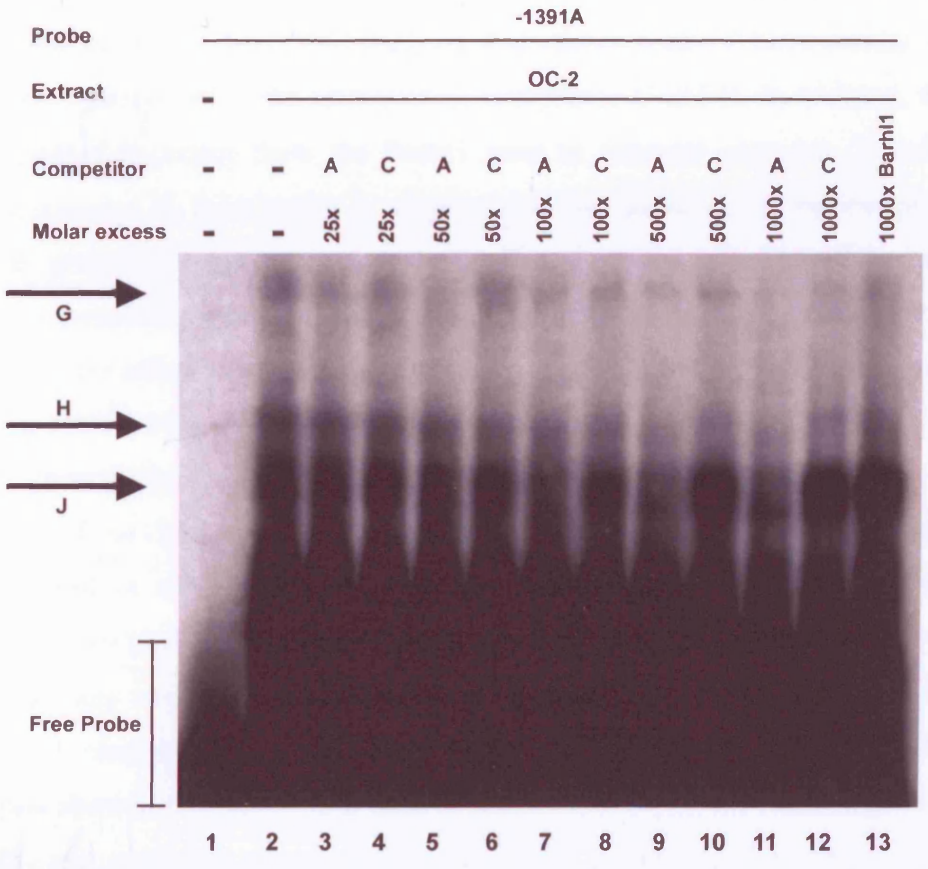


Figure 7.2 Competition analysis to show the relative affinity of alleles -1391A and -1391C for OC-2 derived nuclear proteins. The -1391A probe was incubated in the absence of nuclear extract (lane 1) or in the presence of 8-10μg of OC-2 nuclear extract (lanes 2-13). Increasing amounts of unlabelled A probe (lanes 3, 5, 7, 9 & 11) or unlabelled C probe (lanes 4, 6, 8, 10 & 12) were used to compete with A probe for binding as indicated. An unrelated sequence, a section of exon 1 from the Barhl1 gene was also used as competitor DNA to test for non-specific binding (lane 13). Shifted protein-DNA complexes G, H and J are denoted by the arrows.

probe A (Fig. 7.2, lane 11). However, it is clear that when the unlabelled C probe is used as the competitor, the C allele is unable to induce any significant reduction of complex J, even at 1000-fold competition (Fig. 7.2, comparing lane 1 with 12). These results suggest that allele A has a greater affinity for nuclear protein J compared to allele C, approximately 10-fold greater (comparing relative levels of competition in lane 7 with lane 12, Fig. 7.2) and indicate that the interaction of nuclear protein J with alleles A and C is sequence specific. The failure of 1000-fold excess of an unrelated sequence from the Barhl1 gene to compete nuclear protein J from probe A (Fig. 7.2, lane 13) further confirms the sequence specificity of the interaction between allele A and nuclear protein J.

From the competition experiment presented in Fig. 7.2 it is also evident that complexes G and H are progressively competed from probe A showing a gradual reduction in intensity as the amount of unlabelled excess A or C probe is increased (Fig. 7.2 comparing complex G and H in lanes 3-12 with lane 2). However, the intensity of complexes G and H, although progressively reduced, is similar between both probes at each level of competition. This suggests that alleles A and C have similar affinities for the proteins involved in the formation of complexes G and H. In addition, the ability of an unrelated sequence from the Barhl1 gene to compete complex G and to a lesser extent complex H, from probe A suggests that the interaction of nuclear proteins G and H with probe A is not sequence specific (Fig. 7.2, lane 13). Therefore, in subsequent analysis complex J was the focus of further experiments as the data suggest that allele A has a greater affinity for nuclear protein J compared to allele C and that the affinity of nuclear protein J for allele A is sequence specific.

Initial EMSA analysis suggests that nuclear protein J has a greater affinity for allele A of the -1391 substitution compared to allele C (Fig. 7.2). This observation was investigated in more depth using further EMSA analysis. The ability of probe A to retain nuclear protein J binding in the presence of an excess of either the unlabelled A or C probe was observed and compared to the ability of the C probe to retain nuclear protein J binding in the presence of the same competitors (Fig. 7.3). This line of analysis showed that probe A is able to retain nuclear protein J binding in the presence of 500- and 1000-fold excess of the unlabelled C probe (Fig. 7.3 comparing lane 3 with 6 and 8, respectively). In contrast, in the presence of 500- and 1000-fold excess of the unlabelled A probe there is a clear reduction in nuclear protein J binding (Fig. 7.3 comparing lane 3 with 5 and 7). These results are consistent with the original EMSA observations, Fig. 7.2 suggesting that allele A has a greater affinity for nuclear protein J

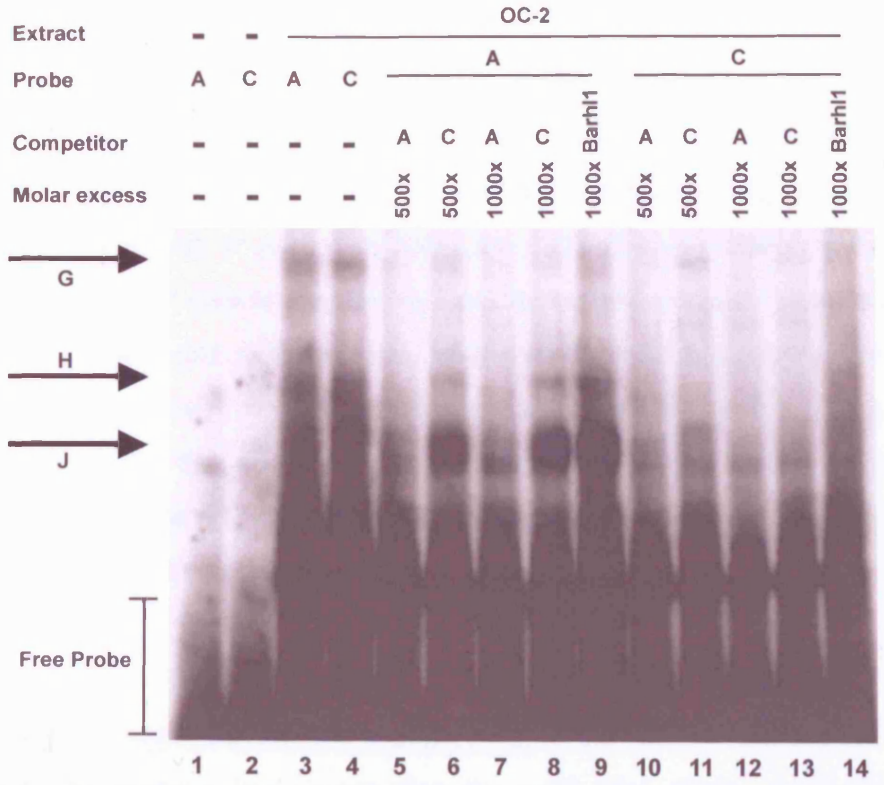


Figure 7.3 The wild-type allele, A, of the -1391 single nucleotide substitution has a greater affinity for OC-2 derived nuclear protein J compared to the variant allele, C. The -1391A probe (lanes 1, 3, 5-9) or -1391C probe (lanes 2, 4, 10-14) were incubated in the absence of nuclear extract (lanes 1-2) or in the presence of 8-10μg of OC-2 nuclear extract (lanes 2-14); in the absence of competitor DNA (lanes 1-4) or with a molar excess of unlabelled: A probe (lanes 5, 7, 10 & 12), C probe (lanes 6, 8, 11 &13) or unrelated sequence, a section of exon 1 of the Barhl1 gene (lanes 9 & 14) as indicated. Shifted protein-DNA complexes G, H and J are denoted by the arrows.

compared to allele C. In comparison, when the C allele sequence is used as the probe and the competition analysis is repeated probe C is unable to strongly retain nuclear protein J binding in the presence of excess unlabelled A or unlabelled C probes (Fig. 7.3 comparing lane 4 with 10-13). In addition, an excess of the unlabelled A probe appears to reduce nuclear protein J binding on the C probe to a greater extent than an equivalent excess of the unlabelled C probe (Fig. 7.3 comparing lanes 10 with 11 and 12 with 13). This is consistent with the data from the reverse experiment (Fig. 7.3 lanes 5-8) and the original observations (Fig. 7.2) suggesting that allele A has a greater affinity for nuclear protein J compared to allele C. Taken together, these observations (Fig. 7.2 and 7.3) confirm that the wild-type allele, A, of the -1391 substitution has a greater affinity for OC-2 derived nuclear protein J compared to the variant allele, C. The ability of 1000-fold excess of an unrelated sequence from the Barhl1 gene to compete with the C probe but not the A probe for nuclear protein J binding (Fig. 7.3 comparing lanes 3 with 9 and 4 with 14) is consistent with these observations and further suggests that the affinity of

nuclear protein J for allele C is very low or that the interaction of allele C with nuclear protein J is non-specific.

In summary, from initial EMSA analysis performed on the -1391 single nucleotide substitution it is evident that a difference in affinity between alleles A and C for nuclear protein J is only apparent if nuclear protein J binding on probe A or probe C is competed for with an excess of either the unlabelled A probe or unlabelled C probe (Fig. 7.2 and 7.3). This is presumably because nuclear protein J is an abundant protein in the OC-2 cell line and therefore, in the absence of competitor DNA both alleles exhibit similar levels of binding. However, competition analysis (Fig. 7.2 and 7.3) clearly illustrates that the -1391 sequence variant in the Brn-3c promoter modifies binding affinity for OC-2 derived nuclear protein J. Therefore, in subsequent analysis investigation was based on identification of nuclear protein J.

7.2.1 Identification of nuclear protein J.

Competition analysis has shown that the wild type allele, A of the -1391 single nucleotide substitution in the Brn-3c promoter has a greater affinity for OC-2 derived nuclear protein J compared to the variant allele, C. Therefore, attempts were made to identify nuclear protein J. MatInspector software (described in section 5.3.1, page 177; www.genomatix.de/products/MatInspector/index.html) was used to screen each allele of the -1391 sequence variant for putative cis-acting elements. The results of this analysis implicated early growth response gene 3 product (Egr-3) and growth factor independence-1 (Gfi-1) as possible candidates for nuclear protein J based on similarity of the sequence in the vicinity of the -1391 sequence variant with the consensus sequences for these transcription factors (see Appendix F). A putative Gfi-1 binding site is located immediately 3' of the -1391 sequence variant; the 5' nucleotide of the core sequence used by MatInspector software to identify putative Gfi-1 binding sites is only 8bp downstream of the -1391 sequence variant. Whereas, a putative Egr-3 binding site is located immediately 5' of the -1391 sequence variant; the 3' nucleotide of the core sequence used by MatInspector software to identify putative Egr-3 binding sites is only 6bp upstream of the -1391 sequence variant (see Appendix F).

Egr-3 and Gfi-1 are both zinc finger transcription factors characterised by three (Patwardhan et al, 1991) and six (Gilks et al, 1993) copies of the $(Cystein)_2(Histidine)_2$ type zinc finger motifs, respectively. Egr-3 belongs to the Egr family of immediate early response genes that are transiently activated in response to environmental signals

includingsignals that induce proliferation, differentiation and apoptosis (for review see Thiel and Cibelli et al, 2002). Gfi-1 is a cellular proto-oncogene that functions as a transcriptional repressor via a 20 amino-acid N-terminal Snail/Gfi-1 (SNAG) domain (Grimes et al, 1996a; Zweidler-Mckay, 1996; for review see Moroy, 2005). Currently, only one other closely related homologue has been identified, Gfi-1b (Tong et al, 1998).

To assess whether nuclear protein J may be a member of the Egr or Gfi-1 transcription factor families competition analysis was performed; the A probe was incubated with OC-2 nuclear protein extracts in the presence of increasing amounts (25-, 100-, 500- and 1000-fold excess) of either the unlabelled Egr consensus sequence (Fig. 7.4, lanes10-13) or the unlabelled Gfi-1 consensus sequence (Fig. 7.4, lanes 5-8; for consensus sequences see Table 3.2 method section 3.2.18). This approach showed that complex J is not competed from probe A by the addition of excess unlabelled Egr consensus sequence; even at 1000-fold excess there is no significant reduction in the intensity of complex J on probe A (Fig. 7.4, comparing lane 2 with 10-13). In contrast, partial competition of nuclear protein J from probe A is evident at 500-fold excess of the unlabelled Gfi-1 consensus sequence (Fig. 7.4 comparing lane 2 with 7). However, this competition does not appear to proceed to completion; even when the amount of unlabelled Gfi-1 consensus sequence as competitor is raised to 1000-fold excess, protein J although further reduced, is not abolished from probe A (Fig. 7.4 comparing lane 2 with 8). Taken together these results imply that nuclear protein J is unlikely to be a member of the Egr transcription factor family but suggest that nuclear protein J may be a member of the Gfi-1 transcription factor family. The failure of an unlabelled oligonucleotide containing a mutated Gfi-1 binding site to induce any significant reduction of nuclear protein J from probe A (Fig. 7.4, comparing lanes 2 and 9) is in agreement with this suggestion as it implies that the interaction of nuclear protein J with the Gfi-1 consensus sequence is sequence specific in that it is dependent on the presence of a functional Gfi-1 binding site. However, compared with previous observations (Fig. 7.2 and 7.3 page 239 and 241, respectively) the data shows that nuclear protein J although having some affinity for the Gfi-1 consensus sequence, has a greater and more convincing affinity for allele A; possibly at least 2-fold more (comparing relative levels of competition in lane 5, Fig. 7.3 page 241 with lane 8, Fig.7.4). If nuclear protein J is Gfi-1 one would generally expect the relative affinity of nuclear protein J for the Gfi-1 consensus binding site to be greater than, or at least equivalent to that exhibited for allele A. In which case it may be that nuclear protein J is not Gfi-1 but instead a different member of this transcription factor family. It should also be considered that

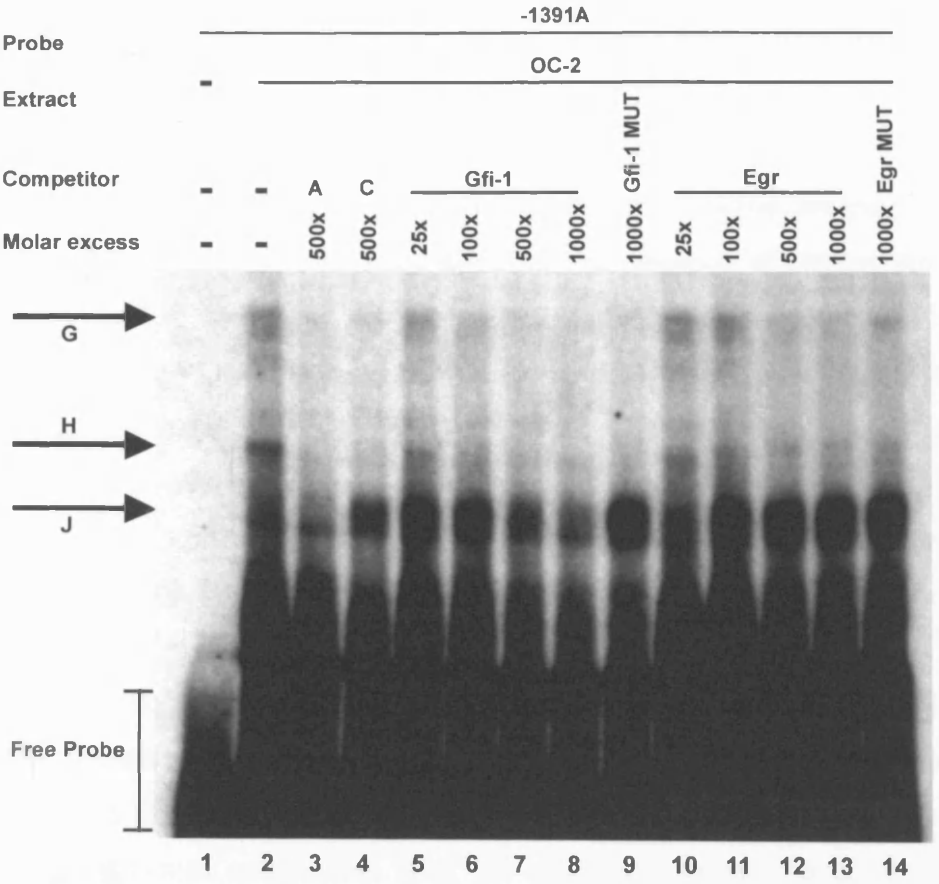


Figure 7.4 Competition analysis using consensus transcription factor binding sites as competitor to identify nuclear protein J. The -1391A probe was incubated in the absence of nuclear extract (lane 1) or in the presence of 8-10μg of OC-2 nuclear extract (lanes 2-14) either in the absence of competitor (lanes 1-2) or with a molar excess of unlabelled: A probe (lane 3), C probe (lanes 4), Gfi-1 consensus sequence (lanes 5-8), an oligonucleotide containing a mutated Gfi-1 binding site (lane 9), Egr consensus sequence (lanes 10-13) or an oligonucleotide containing a mutated Egr binding site (lane 14) as indicated. Shifted protein-DNA complexes G, H and J are denoted by the arrows.

nuclear protein J could be a member of a different transcription factor family, albeit one that exhibits low affinity for the Gfi-1 consensus sequence.

7.2.1.1 Investigating binding of OC-2 derived nuclear proteins to the Gfi-1 consensus sequence.

The data suggest that nuclear protein J could be a member of the Gfi-1 transcription factor family. However, evidence that nuclear protein J is Gfi-1 is not very convincing; competition with 1000-fold excess of the unlabelled Gfi-1 consensus sequence does not completely abolish the interaction of nuclear protein J with probe A (Fig. 7.4 comparing lane 2 with 8). It is possible that this is due to the fact that nuclear protein J is a very abundant protein within the OC-2 cell line, but even so the unlabelled A probe can induce more convincing competition, even when present at a lower molar excess compared to the unlabelled Gfi-1 consensus sequence (comparing relative levels of

competition in Fig. 7.3, lane 5 with Fig. 7.4, lane 8 page 241 and 244, respectively). Therefore, to seek further clarification EMSA analysis was performed using the Gfi-1 consensus sequence as the probe.

OC-2 nuclear extracts were incubated with either the labelled Gfi-1 consensus sequence or probe A in order to compare the pattern of nuclear protein-DNA complex formation between these two sequences. In the absence of competitor DNA this approach revealed shifted nuclear protein-DNA complexes on the Gfi-1 consensus probe that appear analogous to complexes G, H and J on probe A (Fig. 7.5, comparing lanes 1 & 2). Hence, suggesting that both these sequences bind the same protein(s). However, in addition a unique complex termed K is also observed on the Gfi-1 consensus probe that is not observed on probe A (Fig. 7.5 comparing lanes 1 & 2). Competition analysis with an excess of the unlabelled Gfi-1 consensus sequence shows that the unlabelled Gfi-1 consensus sequence can compete with the Gfi-1 consensus probe for binding nuclear protein K (for simplicity one protein, termed K will be assumed to bind) but appears unable to significantly compete for nuclear protein J; complex K is abolished at 500-fold competition in contrast to complex J that still persists at 1000-fold competition (Fig. 7.5 comparing lanes 2, 3 & 4). Moreover, an excess of the unlabelled A probe does not appear to compete with the Gfi-1 consensus probe for binding nuclear protein J; even at 1000-fold competition (Fig. 7.5 comparing lane 2 with 6). These results suggest that the nuclear protein involved in the formation of complex J on probe A may not necessarily be the same one involved in the formation of the analogous complex J on the Gfi-1 consensus probe. Indeed, from previous experiments it is clear that nuclear protein J that binds probe A has a greater affinity for allele A compared to the Gfi-1 consensus sequence (comparing levels of competition in Fig. 7.3, lane 5 with Fig.7.4, lane 8 page 241 and 244, respectively). Hence, if nuclear protein J that binds probe A is the same protein that binds at the equivalent position on the Gfi-1 consensus probe one would generally expect an excess of the unlabelled A probe to compete with the Gfi-1 consensus probe for nuclear protein J.

Interestingly, when the Gfi-1 consensus probe is incubated with OC-2 nuclear protein extract in the presence of an excess of either unlabelled A or C probe, a shifted protein-DNA complex termed SC, can be observed high on the gel (Fig. 7.5 lanes 5, 6, 8 and 9). This complex is not observed when OC-2 nuclear protein extract is incubated with the Gfi-1 consensus probe in the absence of competitor (Fig. 7.5, lane 2) nor if the competitors are changed to the Gfi-1 consensus sequence (Fig. 7.5, lanes 3 & 4) or an oligonucleotide containing a mutated Gfi-1 binding site (Fig. 7.5, lanes 10 & 11). Such

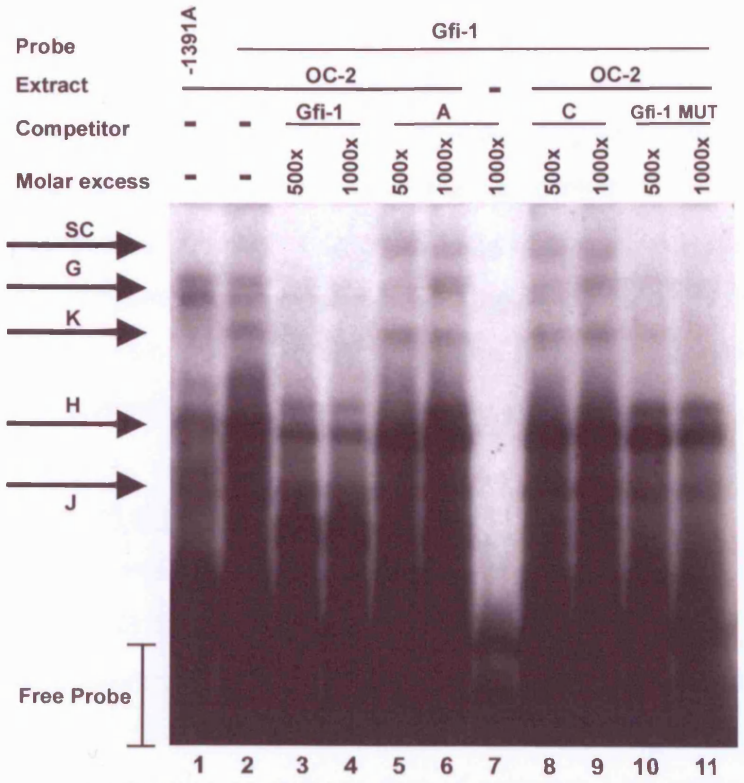


Figure 7.5 Competition analysis using the Gfi-1 consensus sequence as probe. The -1391A probe (lane 1) or labelled Gfi-1 consensus sequence (lanes 2-11) were incubated in the absence of nuclear extract (lane 7) or in the presence of 8-10μg of OC-2 nuclear extract (lanes 1-6 & 8-11) either in the absence of competitor (lanes 1-2) or with a molar excess of unlabelled: Gfi-1 consensus sequence (lanes 3-4), A probe (lanes 5-7), C probe (lanes 8-9) or oligonucleotide containing a mutated Gfi-1 binding site (lanes 10-11) as indicated. Shifted protein-DNA complexes G, H and J are denoted by the arrows. SC denotes a novel shifted protein-DNA complex (observed in lanes 5, 6, 8 & 9)

a complex is difficult to interpret, but it is possible complex SC is a result of a nuclear protein binding to both the Gfi-1 consensus probe and to A or C competitor probes forming one large complex with the reduced mobility as observed. The absence of complex SC when 1000-fold excess of the unlabelled A probe is incubated with the Gfi-1 consensus probe in the absence of OC-2 nuclear extract is consistent with this suggestion (Fig. 7.5 lane 7); formation of complex SC on the Gfi-1 consensus probe is dependent on the presence of OC-2 nuclear extract in addition to an excess of either the unlabelled A or C probes.

In summary, evidence that nuclear protein J that binds probe A is Gfi-1 is not convincing. Indeed, nuclear protein J appears to have only a low affinity for the Gfi-1 consensus sequence (Fig. 7.4 comparing lane 2 with 8, page 244). It is possible this is because nuclear protein J is not Gfi-1 but instead a different member of the Gfi-1 family or a member of a different transcription factor family. In line with these observations it is possible that nuclear protein K, which preferentially binds to the Gfi-1 consensus

sequence compared to allele A may well be Gfi-1. Nuclear protein K is competed from the Gfi-1 consensus probe by 500-fold excess of the unlabelled Gfi-1 consensus sequence but is not competed from the Gfi-1 consensus probe by 1000-fold excess of either the unlabelled A or C probe (Fig. 7.5, comparing lane 2 with 6 & 9, respectively). Hence, suggesting that the interaction of nuclear protein K with the Gfi-1 consensus sequence is sequence specific. However, competition is also evident with 500-fold excess of an oligonucleotide containing a mutated Gfi-1 binding site (Fig. 7.5 comparing lane 2 with 10), which argues against sequence specificity of nuclear protein K with the Gfi-1 consensus sequence.

7.2.1.2 Supershift analysis to identify nuclear protein J.

Initial competition analysis suggests that nuclear protein J may be a member of the Gfi-1 transcription factor family (Fig. 7.4). In an attempt to clarify this data, EMSA analysis was performed using the Gfi-1 consensus sequence as the probe (Fig. 7.5). This approach revealed that although the Gfi-1 consensus sequence can bind a nuclear protein(s) in an equivalent position to that of complex J on probe A, nuclear protein J observed on probe A appears to be a different protein to that observed at the equivalent position on the Gfi-1 consensus probe. In addition, a unique complex, K, is observed on the Gfi-1 consensus probe, but not on probe A raising the possibility that this could be due to Gfi-1 binding. Although, the sequence specificity of nuclear protein K with the Gfi-1 consensus sequence is questionable when tested by competition analysis. Therefore, in an attempt to clarify unequivocally the identity of nuclear proteins J and K supershift analysis was performed with a goat polyclonal anti-Gfi-1 antibody [Gfi-1 (N-20) X, Santa Cruz Biotechnology, Inc.].

The anti-Gfi-1 antibody was incubated with OC-2 nuclear extracts both prior to and subsequent to addition of either probe A or the Gfi-1 consensus probe (Fig. 7.6). In the presence of probe A, this line of analysis failed to either result in the disappearance of nuclear-protein DNA complex J, or induce a supershift of nuclear protein J (Fig. 7.6 comparing lanes 8, 10 and 9, respectively). Similarly, in the presence of the Gfi-1 consensus probe this approach did not result in the disappearance of nuclear-protein DNA complex K nor induce a supershift of nuclear protein K; the nuclear protein that is unique to the Gfi-1 consensus probe (Fig. 7.6 comparing lanes 3, 5 and 4, respectively). This was despite several attempts to optimise the supershift analysis by varying the temperature and the length of the incubation when the anti-Gfi-1 antibody was added prior to probe (4°C overnight or on ice for 30 minutes) and following addition of

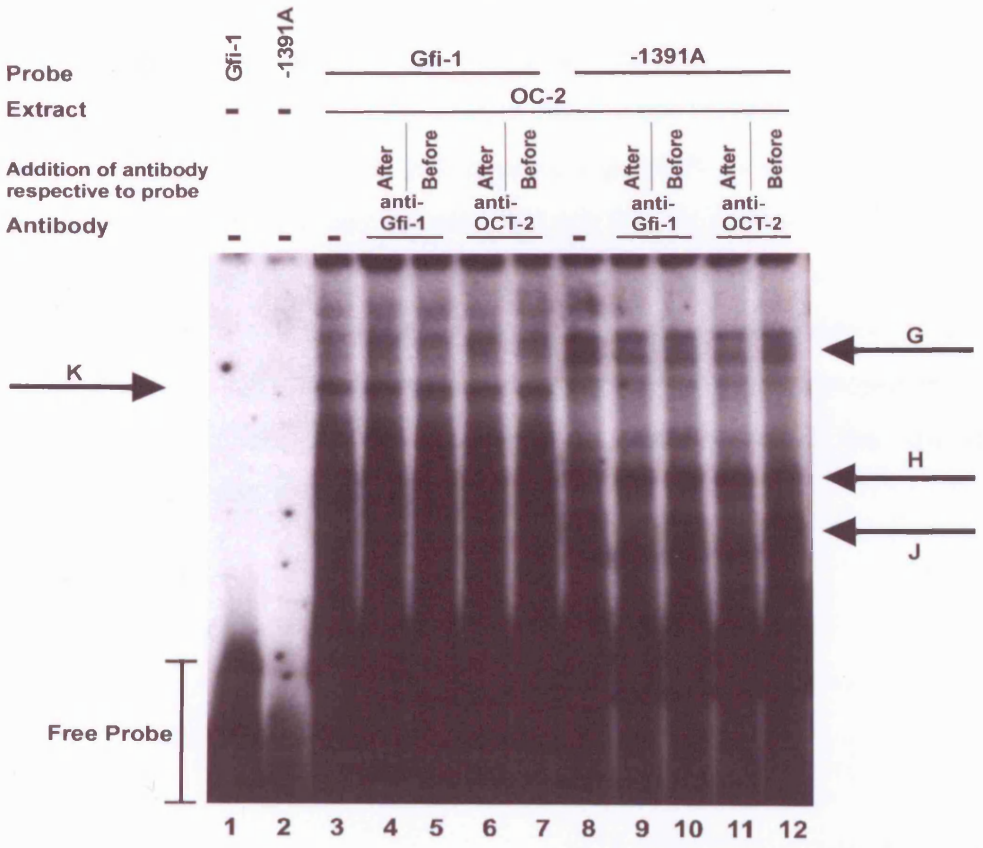


Figure 7.6 Supershift analysis of nuclear proteins J and K. The labelled Gfi-1 consensus sequence or -1391A probe were incubated in the absence of nuclear extract (lanes 1 & 2) or in the presence of 8-10μg OC-2 nuclear extract (lanes 3-12) in the absence of antibody (lanes 3 & 9) or in the presence of anti-Gfi-1 antibody (lanes 4, 5, 9 & 10) or anti-OCT-2 antibody (lanes 6, 7, 11 & 12) as indicated. Antibody was added on ice either 30 minutes prior to probe or subsequent to probe and allowed to incubate for a further 30 minutes on ice as indicated. Protein-DNA complexes G, H, J and K are denoted by the arrows.

labelled probe (on ice or at room temperature for 30 minutes or 1 hour). In addition to attempts to optimise the supershift analysis by varying the amount of anti-Gfi-1 antibody used; either 2, 3 or 4μg of anti-Gfi-1 antibody per binding reaction.

Collectively, these results do not suggest that either nuclear protein J or K is Gfi-1. However, it cannot be categorically excluded on the basis of this supershift analysis that neither nuclear protein J or K are Gfi-1, or a member of the Gfi-1 family. It is possible that the results obtained here are simply due to unsuccessful supershifts. There are many possible reasons to account for the lack of a positive result with the Gfi-1 supershift. Steric interaction between the anti-Gfi-1 antibody and nuclear protein J or K could be prevented if the epitope to which the anti-Gfi-1 antibody was raised is not exposed on either of nuclear proteins J or K and hence, accessible for binding. The presence of other nuclear proteins or the conformation of either nuclear protein J or K under these EMSA conditions could all hinder a positive interaction (see section 5.3.1.3, page 184 for a similar discussion on limitations of supershift analysis). Furthermore, in

the absence of commercially available purified, recombinant Gfi-1 protein it was not possible to test whether the anti-Gfi-1 antibody used in these assay conditions is functional. If the anti-Gfi-1 antibody used in these assay conditions is non-functional this would clearly explain the lack of a positive supershift. In addition, the anti-Gfi-1 antibody used in these assays only detects Gfi-1 not Gfi-1b, the only additional member of the Gfi-1 family that has been identified to date. In which case it is possible that nuclear protein J (or K) could be Gfi-1b. Use of an antibody that detects Gfi-1b would help clarify this issue although due to the time constraints of this project this line of analysis was not possible because functional characterisation of the -3434poly-G polymorphism constituted a large part of the study.

7.3 Effect of the -1391 single nucleotide substitution on Gfi-1 binding.

Supershift analysis with an anti-Gfi-1 antibody did not clarify the identification of nuclear protein J (section 7.2.1.2) and therefore, it remains inconclusive whether nuclear protein J is Gfi-1 or a member of the Gfi-1 family. Consequently, it is unclear whether Gfi-1 binds to the putative Gfi-1 binding site 3' of the -1391 single nucleotide substitution in the Brn-3c promoter (analysis using MatInspector software suggests that the region 3' of the -1391 single nucleotide substitution is a putative Gfi-1 binding site; see Appendix F). The -1391 sequence variant is 8bp upstream of the 5' nucleotide of the core sequence used by MatInspector software to identify putative Gfi-1 binding sites (see Appendix F; see also discussion of analysis using MatInspector software, section 7.2.1, page 242). Hence, although EMSA analysis (section 7.2.1) has failed to clarify the identity of nuclear protein J it remains possible that the -1391 sequence variant could modulate Gfi-1 binding to this putative Gfi-1 binding site in the Brn-3c promoter; protein binding to cis-acting elements can be influenced by adjacent sequence (Dawson et al, 1996a).

The possibility that the -1391 single nucleotide substitution in the Brn-3c promoter could disrupt a functional Gfi-1 binding site is important as recent evidence has shown that Gfi-1 plays an important pro-survival role in hair cells of the auditory system (Wallis et al, 2003). In Gfi-1 mutant mice homozygous for a severe loss of function Gfi-1 allele (due to deletion of a large region of the Gfi-1 coding region including the SNAG domain) hair cells are initially specified in both the vestibular and auditory system but fail to progress along their correct differentiation pathway (Wallis et al, 2003). Specifically, the organisation of hair cells in the vestibular organs and organ of Corti is aberrant. The morphology of the vestibular hair cells is abnormal and outer hair cells of the auditory system exhibit ectopic expression of the neuronal marker TUJ1. Ultimately, a loss of functional Gfi-1 protein in the cochlea results in the progressive degeneration of hair cells. Interestingly, the outer hair cells appear the most susceptible and the degeneration progresses along a basal to apical gradient such that at P14 the organ of Corti is completely destroyed. Typical morphological features of apoptosis have been observed in some of the outer hair cells at E18.5 in the Gfi-1 mutant mice. Leading the proposal that apoptosis may be the mechanism responsible for the hair cell death (Wallis et al, 2003). Clearly, presence of functional Gfi-1 protein is

crucial for survival of cochlear hair cells, at least during development. The developmental expression pattern of Gfi-1 in the inner ear is consistent with this concept; Gfi-1 mRNA and protein are reported to be expressed as early as E12.5 in the prospective sensory epithelia in the otic vesicle and when the sensory organelles are clearly defined, around E18.5, high expression levels of Gfi-1 protein are localised in hair cells (Wallis et al, 2003). Furthermore, recent evidence suggests that Gfi-1 is a downstream target of Brn-3c in hair cells and that the survival of outer hair cells in the cochlea may depend to a large extent on Brn-3c maintaining Gfi-1 expression (Hertzano et al, 2004; this evidence is discussed further in section 7.6 page 264). However, it is not known whether Brn-3c directly regulates Gfi-1, but whether by a direct or indirect mechanism the possibility that Gfi-1 may bind to the Brn-3c promoter particularly in the vicinity of the -1391 sequence variant is important. It is tempting to speculate that Gfi-1 may regulate its own expression by binding to the Brn-3c promoter to form a positive or negative feedback loop. Evidently, the -1391 sequence variant located adjacent to a putative Gfi-1 binding site has the potential to modulate this regulation. Therefore, given the pro-survival role of Gfi-1 in the inner ear and the finding that this gene appears to act downstream of Brn-3c in hair cells experiments were designed to clarify whether Gfi-1 binds to the Brn-3c promoter at the putative Gfi-1 binding site 5' of the -1391 single nucleotide substitution.

7.3.1 Strategy.

It is not known whether the OC-2 cell line actively expresses Gfi-1. The OC-2 cell line was isolated from the H-2Kb-*ts*A58 Immortomouse at E13 immediately before hair cells undergo their final mitosis (Rivolta et al, 1998). It has been established the OC-2 cell line expresses Brn-3c (Rivolta et al, 1998) and therefore it could be reasoned that OC-2 cells also express Gfi-1 based on the evidence suggesting that Gfi-1 is a downstream target of Brn-3c (Hertzano et al, 2004). Indeed, Gfi-1 protein is reported to be expressed as early as E12.5 in the prospective sensory epithelia in the otic vesicle (Wallis et al, 2003). However, semi-quantitative RT-PCR on RNA extracted from cochlea sensory epithelia at several developmental time points did not detect Gfi-1 expression until E15.5 (Hertzano et al, 2004). Moreover, expression of Gfi-1 in hair cells of the inner ear does not appear to be particularly high until the later developmental stages around E18.5 (Wallis et al, 2003; Hertzano et al, 2004). Hence, it

is possible that even if Gfi-1 is actively expressed in the OC-2 cell line the levels of Gfi-1 protein could be very low.

In the first instance attempts were made to characterise endogenous Gfi-1 expression in OC-2 cells by performing western immunoblot analysis on OC-2 nuclear extracts using the goat polyclonal anti-Gfi-1 antibody [Gfi-1 (N-20) X, Santa Cruz Biotechnology, Inc.] used in supershift analysis (section 7.2.1.2). It was reasoned that if the OC-2 cell line is confirmed as expressing Gfi-1, then use of more concentrated nuclear extracts in EMSA analysis may help to clarify whether Gfi-1 binds to the region 3' of the -1391 sequence variant. However, this approach was not successful; a single band equivalent to the 55 kDa polypeptide for Gfi-1 was not observed with many additional bands of higher and lower molecular weight also present presumably due to cross-reactivity of the anti-Gfi-1 antibody used (data not shown). In the absence of informative data regarding endogenous expression of Gfi-1 in OC-2 cells and a lack of commercially available purified recombinant Gfi-1 protein an alternative approach was sought to assess whether the region 3' of the -1391 sequence variant constitutes a Gfi-1 binding site; EMSA analysis was performed using OC-2 nuclear extracts that had been prepared from OC-2 cells transfected with an expression vector for Gfi-1.

7.3.2 EMSA analysis using nuclear extracts prepared from OC-2 cells transfected with an expression vector for Gfi-1.

The mammalian expression vector containing the rat cDNA for Gfi-1 subcloned downstream of the cytomegalovirus 5 (CMV5) promoter, termed CMV5-Gfi-1, has been described previously (Grimes et al, 1996b) and was kindly provided by Professor Horwitz (University of Washington, U.S.A). Briefly, OC-2 cells were plated into 90mm dishes (Nunclon™) and transiently transfected with either the CMV5-Gfi-1 expression vector or the empty pCi expression vector (Promega), which is under the control of the cytomegalovirus, CMV promoter. Cells were grown to 80% confluent and harvested for nuclear extract as described (see method section 3.2.4.3). This, in combination with the standard OC-2 nuclear extracts (see method section 3.2.4.2) formed three different types of OC-2 nuclear extract: standard (OC-2 cells not transfected with any expression vector), +CMV5-Gfi-1 (OC-2 cells transfected with the CMV5-Gfi-1 expression vector) and +pCi (OC-2 cells transfected with the empty pCi expression vector).

To assess whether the -1391A >C single nucleotide substitution modulates the binding of Gfi-1 to this region of the Brn-3c promoter the A probe and C probe were

incubated with OC-2 nuclear extracts that had been transfected with either Gfi-1 expression vector (+CMV5-Gfi-1) or empty expression vector (+pCi) as a control for aberrant effects induced by the *CMV* backbone in the empty pCi vector (Fig. 7.7). However, this approach did not show a difference in protein-DNA complex formation between Gfi-1 transfected (+CMV5-Gfi-1) and non-transfected (+pCi) extracts on either probe A or C. (Fig. 7.7 comparing lanes 5 with 6 and 8 with 9). In addition, the same result is observed when the probe is changed to the Gfi-1 consensus sequence as a positive control for binding of Gfi-1 under these EMSA conditions (Fig. 7.7 comparing lane 2 with 3). It is possible that these results are due to a lack of Gfi-1 protein for example, due to either little or no expression from the CMV5-Gfi-1 expression vector. However, judging from the quality of the shifted protein-DNA complexes formed with: the standard OC-2 nuclear extracts, nuclear extracts that have been transfected with the CMV5-Gfi-1 expression vector (+CMV5-Gfi-1) or empty pCi expression vector (+pCi) it appears more conceivable that the results are due to poor quality nuclear extracts. The intensity of complexes J and G on probes A and C is faint compared to previous EMSA analysis (comparing complex J and G in Fig. 7.7, lanes 4-9 with for example, Fig. 7.5,

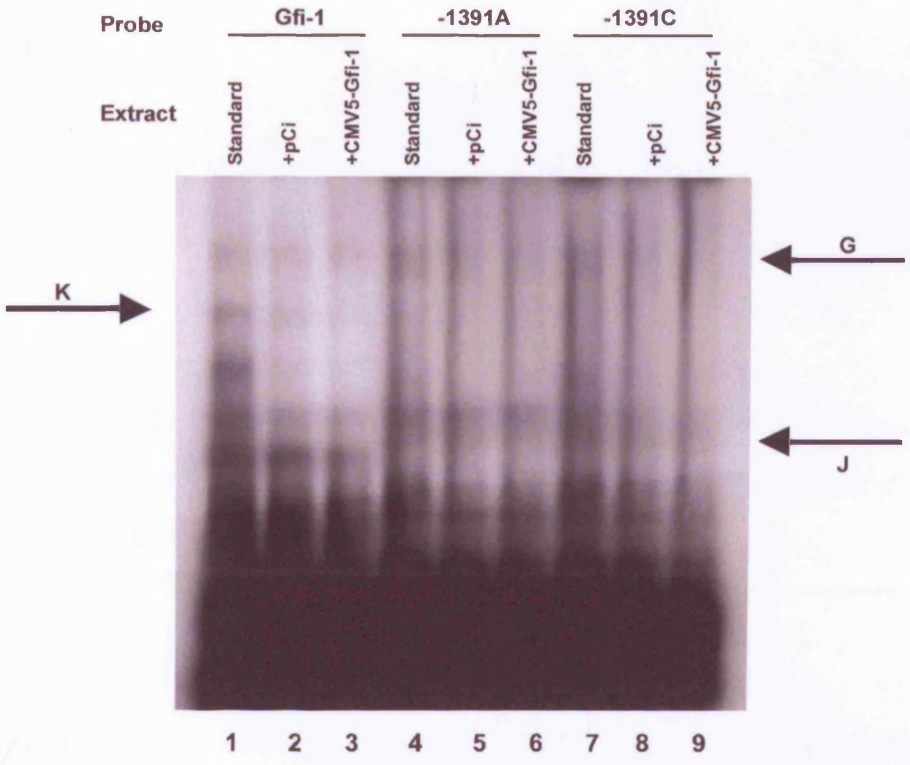


Figure 7.7 Effect of the -1391 single nucleotide substitution on Gfi-1 binding. The labelled Gfi-1 consensus sequence (lanes 1-3), -1391A probe (lanes 4-6) or -1391C probe (lanes 7-9) were incubated with 8-10μg of: standard OC-2 nuclear extract (lanes 1, 4 & 7), OC-2 nuclear extract that had been transfected with the empty pCi expression vector, which is under the control of the CMV promoter (lanes 2, 5 & 8) or OC-2 nuclear extract that had been transfected with the CMV5-Gfi-1 expression vector (lanes 3, 6 & 9). Protein-DNA complexes G, J and K are denoted by the arrows.

lane 1, page 246 or Fig. 7.6, lane 8, page 248) and complex H cannot be observed (comparing lanes 4-9 in Fig. 7.7 with for example lane 1 in Fig. 7.5, page 246 or lane 8 in Fig. 7.6, page 248). A western immunoblot of OC-2 nuclear extracts that have been transfected with the CMV5-Gfi-1 expression vector (+CMV5-Gfi-1) may clarify if these results are due to a lack of Gfi-1 protein. However, this approach has been tried previously on the standard OC-2 nuclear extracts in an attempt to characterise endogenous Gfi-1 expression and was not successful due to apparent cross-reactivity of the anti-Gfi-1 antibody used. Therefore, western immunoblot analysis of OC-2 nuclear extracts that have been transfected with the CMV5-Gfi-1 expression vector (+CMV5-Gfi-1) was not undertaken. Had time permitted repeating and optimising this experiment with more concentrated batches of OC-2 nuclear extracts may have given more informative results as to whether the region 5' of the -1391 sequence variant constitutes a Gfi-1 binding site, but it wasn't possible to pursue this line of analysis as the main focus of investigation was centred around the -3434 poly-G polymorphism (see Chapter 5).

7.4 Effect of the -1391A >C single nucleotide substitution on basal Brn-3c promoter activity and activity of the Brn-3c promoter in response to exogenous Gfi-1.

It has been established that the single nucleotide substitution at position -1391 in the Brn-3c promoter modifies binding affinity for a specific nuclear protein (protein J) expressed in OC-2 cells. From EMSA competition analysis it is clear that the wild-type allele, A, of the -1391 sequence variant has a higher affinity for nuclear protein J compared to the variant allele, C. Using MatInspector software (see section 7.2.1) two putative cis-acting elements in the vicinity of the -1391 sequence variant: Egr-3 and Gfi-1 were identified and attempts were made to identify nuclear protein J. Competition analysis ruled unambiguously that Egr or a member of this transcription factor family is unlikely to be nuclear protein J and in contrast, suggested that nuclear protein J may be a member of the Gfi-1 transcription factor family. Further analysis to clarify whether nuclear protein J is Gfi-1 was not successful but it could not be ruled out unequivocally that nuclear protein J is Gfi-1, or a member of this transcription factor family.

The determination of sequence specific protein binding within the promoter of a gene does not mean that the protein binding site in question is a bona fide functional cis-acting element. Therefore, it needs to be clarified whether the difference in affinity between alleles A and C for nuclear protein J actually has a functional affect on activity of the Brn-3c promoter. It is possible that the basal activity of the Brn-3c promoter is modulated by the difference in affinity between A and C alleles for nuclear protein J. Furthermore, regardless of whether nuclear protein J is a member of the Gfi-1 family, the -1391 sequence variant may modulate a Gfi-1 binding site and this could have a functional affect on activity of the Brn-3c promoter. This is particularly important in light of the recent evidence that Gfi-1 is required for differentiation and survival of inner ear hair cells (Wallis et al, 2003) and the finding that Gfi-1 appears to be a downstream target of Brn-3c in hair cells (Hertzano et al, 2004). Attempts to determine whether Gfi-1 can bind to the putative Gfi-1 binding site 5' of the -1391 sequence variant using nuclear extracts prepared from OC-2 cells that were transfected with an expression vector for Gfi-1 (+CMV5-Gfi-1) were not conclusive due to presumably poor quality nuclear extracts (see Fig 7.7, page 253). Therefore, in an attempt to clarify whether the -1391 sequence variant modulates basal activity of the Brn-3c promoter and whether the region 5' of the -1391 sequence variant is responsive to Gfi-1 transient

transfections were performed in OC-2 cells. Two Brn-3c promoter-luciferase reporter gene constructs were generated that contained either allele A or C at position -1391 and the activity of these constructs was observed at basal levels and in response to an expression vector for Gfi-1 (CMV5-Gfi-1).

The firefly luciferase reporter vector pGL3basic (Promega) carrying 3.6Kb (bases -80bp to -3670bp) of the human Brn-3c promoter (termed: pgl3b-Brn-3c-3.6) had already been cloned in our laboratory from the human PAC library RPCI1 (HGMP, MRC, U.K.). The genetic background of this construct at loci in the Brn-3c promoter established from the PCR-SSCP screen to be polymorphic was determined by direct sequencing and found to be: -3495$(GT)_{15}$, -3457$(GA)_2$, -3432$(G)_8$CG, -1391C, -566$(GT)_{20}$ and -386C (see Table 3.4, method section 3.2.18). Hence, pgl3b-Brn-3c-3.6 carries the variant allele, C at position -1391. Therefore, construct pgl3b-Brn-3c-3.6 (for simplicity hereafter referred to as p-Luc-1391C) was used as a template in a site-directed mutagensis (SDM) reaction (QuikChange®, Stratagene) with primers -1391aSDM-S: 5' GCAGCGTAGTCGAGGTCCAGGATTC 3' and -1391tSDM-AS: 5' GAATCCTGGACCTCGACTACGCTGC 3' to modify the variant allele C at position -1391 to the wild type allele, A, to form construct p-Luc-1391A. This generated two different Brn-3c luciferase-reporter gene constructs that differed in the nature of the -1391 allele; see Fig. 7.8a (see method section 3.2.8.1 for details in construct formation; for convenience in Chapter 6 construct p-Luc-1391A is referred to as p-Luc-SNPG1.$_{566GT20}$).

To assess whether Gfi-1 has an affect on activity of the Brn-3c promoter and whether this response, if any is modulated by the -1391 sequence variant the Gfi-1 expression vector, CMV5-Gfi-1, was co-transfected into OC-2 cells with Brn-3c promoter-luciferase reporter gene constructs that carry either allele A (p-Luc-1391A) or allele C (p-Luc-1391C) at position -1391 (Fig. 7.8b). The results obtained were compared with the respective basal levels of activity for each of these Brn-3c-luciferase reporter gene constructs in the presence of the empty pCi expression vector (Promega) as this vector is under the control of the CMV promoter (Fig. 7.8b). This ensured that the DNA concentration between wells in the absence of the Gfi-1 expression vector was standardised to minimise any possible effect of inter-well variation in DNA concentration on transfection efficiency. Similar co-transfections in the absence and presence of the Gfi-1 expression vector, CMV5-Gfi-1, were performed with the empty pGL3basic vector as a control for activity of Gfi-1 on the backbone of the pGL3basic vector. In these experiments the *renilla* luciferase expression vector pRL-SV40

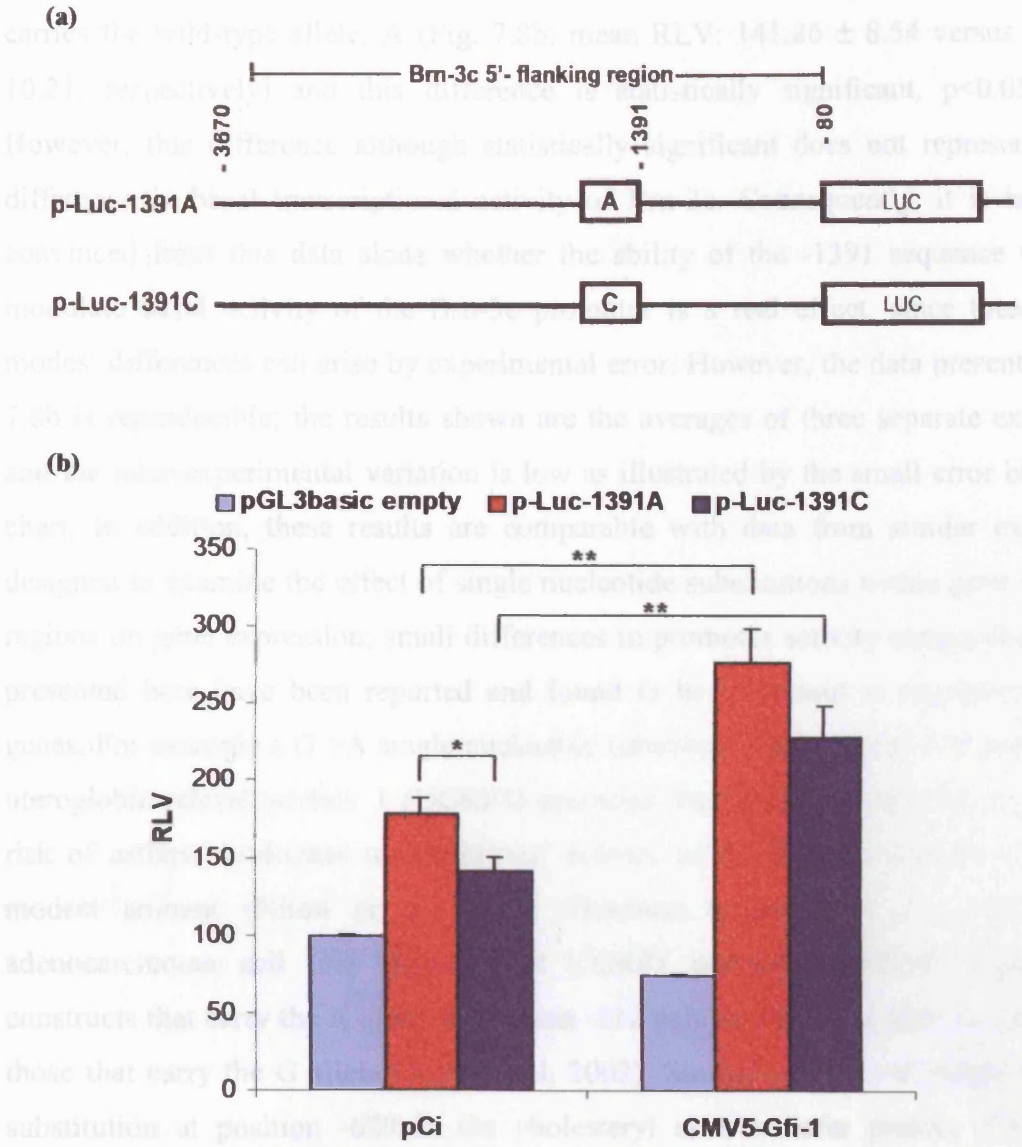


Figure 7.8 Effect of the -1391 single nucleotide substitution on basal Brn-3c promoter activity and activity of the Brn-3c promoter in response to exogenous Gfi-1. (a) A schematic diagram (not to scale) to illustrate 3.6Kb Brn-3c promoter-luciferase reporter gene constructs that contain either the wild-type allele, A or the variant allele, C at position -1391. **(b)** OC-2 cells were co-transfected with 500ng of p-Luc-1391A, p-Luc-1391C or the empty pGL3Basic vector in the presence of 500ng of the Gfi-1 expression vector (CMV5-Gfi-1) or the empty pCi expression vector, which is under the control of the CMV promoter. The *renilla* expression vector pRL-SV40 (100ng) was used as an internal control to normalise for differences in transfection efficiency. Error bars on chart represent the standard error of the mean. RLV: relative luciferase value. Luciferase activities shown on chart are expressed relative to the empty pGL3Basic vector in the presence of pCi expression vector. The experiment was performed in triplicate and repeated at least twice with two different preparations of DNA. * $p<0.05$ (t-test; d.f. = 16), ** $p<0.05$ (paired t-test; d.f. = 1).

(Promega) was used as an internal control to adjust for intra- and inter- experimental variation

This approach showed that in the absence of exogenous Gfi-1, basal activity between Brn-3c promoter-luciferase reporter gene constructs that differ in the nature of the -1391 allele is different. Basal activity from construct p-Luc-1391C that carries the

variant allele, C at position -1391 is 26% less than that from construct p-Luc-1391A that carries the wild-type allele, A (Fig. 7.8b; mean RLV: 141.86 ± 8.54 versus 178.23 ± 10.21, respectively) and this difference is statistically significant, $p<0.05$ (t-test). However, this difference although statistically significant does not represent a large difference in basal transcriptional activity of Brn-3c. Consequently, it is hard to be convinced from this data alone whether the ability of the -1391 sequence variant to modulate basal activity of the Brn-3c promoter is a real effect, since these kind of modest differences can arise by experimental error. However, the data presented in Fig. 7.8b is reproducible; the results shown are the averages of three separate experiments and the inter-experimental variation is low as illustrated by the small error bars on the chart. In addition, these results are comparable with data from similar experiments designed to examine the effect of single nucleotide substitutions within gene regulatory regions on gene expression; small differences in promoter activity comparable to those presented here have been reported and found to be important in regulation of other genes. For example a G >A single nucleotide substitution at position -112 in the human uteroglobin-related protein 1 (UGRP1) promoter that is associated with an increased risk of asthma modulates transcriptional activity of the UGRP1 promoter by only a modest amount (Niimi et al, 2002). Transient transfections in a human lung adenocarcinoma cell line showed that UGRP1 promoter-luciferase reporter gene constructs that carry the A allele at position -112 exhibit 24 % less activity compared to those that carry the G allele (Niimi et al, 2002). Similarly, an A >C single nucleotide substitution at position -629 in the cholesteryl ester transfer protein (CETP) gene promoter has been found to have a functional affect on activity of the CETP promoter such that presence of the A allele reduces promoter activity by 25 % compared to the C allele (Dachet et al, 2000). Moreover, the reduced basal activity observed with construct p-Luc-1391C that carries the variant allele, C at position -1391 is consistent with EMSA analysis; the variant allele C has a reduced affinity for OC-2 derived nuclear protein J compared to the wild-type allele, A (see section 7.2, Fig. 7.2 and 7.3 page 239 and 241, respectively). Furthermore, if quantitative differences in Brn-3c expression are a risk factor for late onset hearing loss it is likely that one may only find subtle changes in Brn-3c promoter activity. Indeed, it must be remembered that late onset hearing loss is a complex trait; multiple loci are thought to underlie susceptibility and variation in gene expression or function at any one locus is thought to be moderate (for review see Fransen et al, 2003).

In the presence of an expression vector for Gfi-1 transient transfection assays showed that activity from construct p-Luc-1391A that carries the wild-type A allele and construct p-Luc-1391C that carries the variant C allele at position -1391 increases by 56% and 61% respectively, compared to basal levels (Fig. 7.8b). This increase in Brn-3c promoter activity although moderate, is statistically significant, $p<0.05$ (paired t-test; Fig. 7.8b). The absence of any increased activity from the empty pGL3basic vector when co-transfected with the Gfi-1 expression vector confirms that this response is not due to the backbone of the pGL3basic vector but specific to the Brn-3c promoter insert. Hence, it is clear that regardless of whether the wild type A allele or variant C allele is present at position -1391 Gfi-1 can transactivate the Brn-3c promoter (Fig. 7.8b). This increase in Brn-3c promoter activity does not differ significantly between Brn-3c promoter-luciferase reporter gene constructs that contain either allele A or C at position -1391. Thus, suggesting that the ability of the Brn-3c promoter to respond to Gfi-1 is independent of the nature of the allele at position -1391. This could be explained if Gfi-1 does not bind to the putative Gfi-1 binding site 3' of the -1391 sequence variant or if Gfi-1 bound to this region 3'of the -1391 sequence variant but the single nucleotide substitution had no, or minimal, effect on the ability of Gfi-1 to bind. This is possible as the -1391 sequence variant is not located in a core residue of the Gfi-1 consensus binding site (see Appendix F). Further investigation is needed to clarify this issue. It is possible that when Gfi-1 levels are limiting the -1391 sequence variant has an affect on Brn-3c promoter activity; this may account for the subtle, but significant difference in activity observed under basal conditions (Fig. 7.8b). However, it is evident from this analysis that whether through a direct or indirect mechanism Gfi-1 can increase transcriptional activity of the Brn-3c promoter (Fig. 7.8b). This finding also lends support to the proposal that the absence of a band indicative of Gfi-1 protein binding when the Gfi-1 consensus probe is incubated with nuclear extracts that have been transfected with an expression vector for Gfi-1, +CMV5-Gfi-1, (Fig. 7.7, page 253) is due to poor quality or insufficient quantity of nuclear extract and not absence of Gfi-1 expression from the CMV5-Gfi-1 expression vector.

7.5 Effect of the -1391 single nucleotide substitution in the context of a heterologous promoter.

Transient transfection assays using luciferase reporter gene constructs carrying a 3.6Kb Brn-3c promoter fragment have shown that Gfi-1 can transactivate the Brn-3c promoter (Fig. 7.8b, page 257). In these experiments the nature of the allele at position -1391 did not significantly modulate the response of the Brn-3c promoter to Gfi-1. It therefore still remains unclear whether Gfi-1 does actually bind to the putative Gfi-1 binding site 3' of the -1391 sequence variant; both EMSA analysis and reporter gene assays have failed to clarify whether this is the case. Hence, it is unclear whether the putative Gfi-1 binding site 3' of the -1391 sequence variant is responsible for the response of the Brn-3c promoter to Gfi-1.

In an attempt to establish whether the putative Gfi-1 binding site 3' of the -1391 sequence variant is responsive to Gfi-1 the *isolated* wild-type allele, A, and variant allele, C, were examined in the context of the SV40 heterologous promoter. The pGL3-Promoter vector (Promega) contains an SV40 promoter cloned upstream of the firefly luciferase gene (*Fluc*+) and has been discussed previously (see section 5.7, page 213). Using the synthetic oligonucleotides that were devised for EMSA analysis of the -1391 sequence variant (see Table 3.2 method section 3.2.18) a series of pGL3-Promoter constructs were devised that contained two or three concatenated copies of: allele A [pGL3-P(1391A)$_2$], allele C [pGL3-P(1391C)$_2$], the consensus sequence for Gfi-1 [pGL3-P(Gfi-1)$_2$] or an oligonucleotide containing a mutated Gfi-1 binding site [pGL3-P(Gfi-1Mut)$_3$] cloned upstream of the heterologous SV40 promoter (Fig. 7.9a). These constructs were co-transfected into OC-2 cells with an expression vector for Gfi-1 (CMV5-Gfi-1) and activity was compared to basal levels in the presence of the empty pCi expression vector as this is under the control of the CMV promoter (Fig. 7.9b). In these experiments the pRL-SV40 vector (Promega) was used as an internal control to adjust for intra- and inter- experimental variation.

This line of analysis showed that in the presence of CMV5-Gfi-1 expression vector activity from the empty pGL3-Promoter vector increases 14-fold (Fig. 7.9b). This suggests that the backbone of the pGL3-Promoter vector is subject to regulation by exogenous Gfi-1. Hence, the validity of the remaining results obtained in the presence of CMV5-Gfi-1 expression vector is confounded and valid conclusions cannot be drawn. In particular, it is not possible to ascertain whether the putative Gfi-1 binding

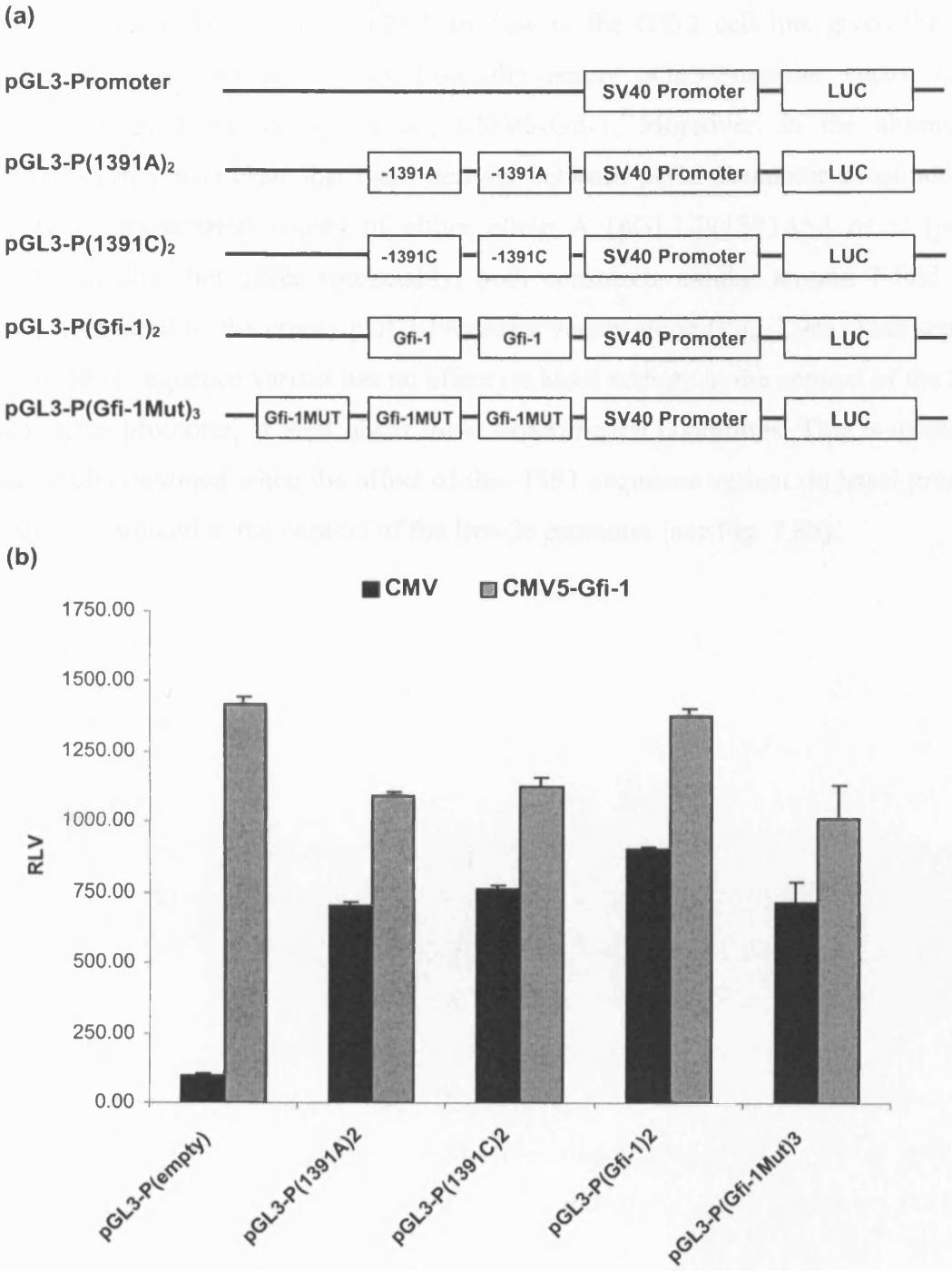


Figure 7.9 Effect of the -1391A >C single nucleotide substitution in the context of a heterologous promoter. (a) Schematic diagram (not to scale) to illustrate pGL3-Promoter luciferase reporter gene constructs that differ in the nature of the -1391 allele. Constructs contain two or three concatenated copies of either allele A [pGL3-P(1391A)$_2$], allele C [pGL3-P(1391C)$_2$], the consensus sequence for Gfi-1 [pGL3-P(Gfi-1)$_2$] or an oligonucleotide containing a mutated Gfi-1 binding site [pGL3-P(Gfi-1 Mut)$_3$] cloned upstream of the SV40 promoter in pGL3-Promoter vector as indicated. **(b)** OC-2 cells were co-transfected with 500ng of either pGL3-P(empty), pGL3-P(1391A)$_2$, pGL3-P(1391C)$_2$, pGL3-P(Gfi-1)$_2$ or pGL3-P(Gfi-1 Mut)$_3$ in the presence of 500ng of either Gfi-1 expression vector (CMV5-Gfi-1) or the empty pCi expression vector which is under the control of the CMV promoter. The *renilla* expression vector pRL-SV40 (100ng) was used as an internal control to normalise for differences in transfection efficiency. Error bars on chart represent the standard error of the mean. RLV: relative luciferase value. Luciferase activities shown on chart are expressed relative to pGL3-P(empty) in the presence of pCi expression vector by assigning a relative value of 100 (RLV). The experiment was performed in triplicate and repeated at least twice with two different preparations of DNA.

site 3' of the -1391 sequence variant is responsive to Gfi-1. However, from this analysis it can be inferred that levels of Gfi-1 are low in the OC-2 cell line given the large increase in transcriptional activity from the empty pGL3-Promoter vector in the presence of Gfi-1 expression vector, CMV5-Gfi-1. Moreover, in the absence of exogenous Gfi-1 it is clear that basal activity between pGL3-Promoter constructs that carry two concatenated copies of either allele A [pGL3-P(1391A)$_2$] or C [pGL3-P(1391C)$_2$], does not differ appreciably; both constructs exhibit around 7-fold more activity compared to the empty pGL3-Promoter vector alone (Fig. 7.9b). This suggests that the -1391 sequence variant has no affect on basal activity in the context of the SV40 heterologous promoter, at least under these experimental conditions. This is in contrast to the results obtained when the affect of the -1391 sequence variant on basal promoter activity is examined in the context of the Brn-3c promoter (see Fig. 7.8b).

7.6 General discussion.

Gfi-1 was originally characterised as a transcriptional repressor (Grimes et al, 1996a; Zweidler-McKay et al, 1996) and the majority of the published data is in agreement with this concept (Grimes et al, 1996b; Duan and Horwitz, 2003; McGhee et al, 2003). There appears little evidence in the literature for a role of Gfi-1 as a direct activator of transcription; Gfi-1 has only been reported to act as a direct transcriptional activator in the context of an artificial promoter containing four copies of the Gfi-1 consensus sequence in an erythroid cell line (Osawa et al, 2002). However, it is conceivable that Gfi-1 can function as a dual transcriptional repressor and activator of transcription. Binding of some trans-acting factors to their cognate sites has a positive or negative affect on transcription depending on the cell type and promoter context as co-factors needed for repression and / or activation can vary between cell type and flanking DNA sequences can greatly influence the context of transcription factor binding sites. Certainly, there is evidence that the Gfi-1 homologue *senseless* in *Drosophilia* has a dual role as a transcriptional repressor and activator (for review see Jafar-Nejad and Bellen, 2004). Hence, the data presented in this thesis that Gfi-1 can transactivate the Brn-3c promoter is interesting although the extent of transcriptional activation is moderate and further investigation is required to establish whether this is via a direct mechanism. Repeating the EMSA analysis as described in section 7.3.2 with *in-vitro* translated Gfi-1 protein may help clarify whether Gfi-1 can bind to the putative Gfi-1 binding site 3' of the -1391 single nucleotide substitution. This data would help elucidate whether the putative Gfi-1 binding site 3' of the -1391 sequence variant is directly responsible for the response of the Brn-3c promoter to Gfi-1. If it is established that the putative Gfi-1 binding site 3' of the -1391 sequence variant does not bind Gfi-1 foot-printing analysis with a source of exogenous Gfi-1 protein may help determine whether Gfi-1 binds to an alternative region(s) within the 3.6Kb Brn-3c promoter fragment used for transient transfection assays. In any case, subsequent transient transfection assays incorporating site-directed muatagensis studies of the region(s) of the Brn-3c promoter shown to bind Gfi-1 would help clarify unequivocally the region of the Brn-3c promoter directly responsible for the observed response to Gfi-1. In addition, this line of analysis would allow one to establish the key residues of the Brn-3c promoter crucial for Gfi-1 binding.

Certainly, it is conceivable that Gfi-1 positively regulates Brn-3c. It has been reported that Gfi-1 is a downstream target of Brn-3c in hair cells (Hertzano et al, 2004). Consequently, it is tempting to speculate that Gfi-1 regulates its own expression by binding to the Brn-3c promoter to form a positive feedback loop, in effect, promoting its own expression. However, it must be emphasized that the evidence Gfi-1 is a downstream target of Brn-3c in hair cells is not direct. Brn-3c has not been shown to bind to and have a functional affect on the Gfi-1 promoter, but rather by using a comparative approach mRNA expression profiles from E16.5 inner ears of wild-type and Brn-3c mutant mice have shown that a deficiency in Brn-3c leads to a significant reduction in Gfi-1 expression levels (Hertzano et al, 2004). Clearly, whether Brn-3c directly regulates Gfi-1 awaits further investigation, but it seems likely that Gfi-1 acts downstream of Brn-3c in inner ear sensory hair cells. This is of great interest as one role of Gfi-1 at least in other cell types, is to prevent apoptosis. Over-expression of Gfi-1 in immortalized IL-2-dependent T cells and in primary thymocytes results in the down-regulation of the pro-apoptotic genes Bax and Bak and in the inhibition of cell death (Grimes et al, 1996b). Furthermore, as discussed by Hertzano et al, Gfi-1 can promote STAT3 signals by a virtue of its ability to sequester the STAT3 inhibitor PIAS3 (Rodel et al, 2000), and STAT3 signalling has been reported to transactivate members of the anti-apoptotic Bcl-2 family (Fukada et al, 1996; Catlett-Falcone et al, 1999). Therefore, based on this evidence it is tempting to speculate that Gfi-1 may employ similar survival mechanisms in cochlea hair cells of the inner ear.

7.7 Conclusion.

In conclusion, transient transfection assays in OC-2 cells with Brn-3c promoter-luciferase reporter gene constructs suggest that the -1391 single nucleotide substitution modulates basal activity of the Brn-3c promoter; presence of the variant C allele reduces basal transcriptional activity of Brn-3c by 26 % compared to the wild type A allele and this difference is statistically significant $p<0.05$ (t-test; Fig. 7.8b). Importantly, these results are consistent with EMSA analysis designed to examine the affect of the -1391 sequence variant on binding of OC-2 derived nuclear proteins; the variant C allele has a reduced affinity for OC-2 derived nuclear protein J compared to the wild-type A allele (section 7.2). The agreement between the affinity of nuclear protein J for the -1391 allele sequence and the extent of basal transcriptional activity supports the proposal that the observed differences in basal activity of Brn-3c are due to genuine effects induced by the nature of the -1391 allele and collectively, these data suggest that the -1391 single nucleotide substitution in the Brn-3c promoter could have a functional effect. However, it is important to remember that these observations are based within an *in-vitro* context. The significance of these results in terms of basal transcriptional activity of Brn-3c *in-vivo* is not known, but it is possible that the -1391C allele may be associated with decreased basal activity of Brn-3c in inner ear hair cells. Indeed, it is tempting to speculate that individuals who carry the variant C allele could have reduced expression of Brn-3c a hair cell pro-survival gene.

Attempts were made to identify nuclear protein J (see section 7.2.1). The results of this analysis although unable to confirm the identity of nuclear protein J, could not rule out that nuclear protein J is a member of the Gfi-1 transcription factor family. In light of these observations and the recent evidence that Gfi-1 plays an important pro-survival role in cochlea hair cells (Wallis et al, 2003; Hertzano et al, 2004) experiments were devised to establish whether the putative Gfi-1 binding site 3' of the -1391 sequence variant constitutes a functional Gfi-1 binding site (see section 7.3 and 7.4). This line of investigation did not clarify whether Gfi-1 could bind to the putative Gfi-1 binding site 3' of the -1391 sequence variant (see section 7.3) but, it is evident that Gfi-1 can increase transcriptional activity of the Brn-3c promoter in OC-2 cells (see Fig. 7.8b). This finding that Gfi-1 can transactivate the Brn-3c promoter, albeit to moderate extent, is interesting; it raises the possibility that Gfi-1 a well-characterised

transcriptional repressor may function as a dual transcriptional repressor and activator of transcription.

8.0 A Preliminary Association Study To Assess Whether Common Sequence Variants At The Brn-3c Locus Are A Risk Factor For Late Onset Hearing Loss.

8.1 Introduction.

Delineating the genetic basis of late onset hearing loss is a difficult undertaking. Late onset hearing loss is an extremely complex disease genetic susceptibility is confounded by environmental factors including acoustic trauma and ototoxic drugs and this is exacerbated by the polygenic nature of the disease; a complex genetic-environmental interplay is thought to underlie the aetiology (for reviews see Jennings and Jones, 2001; Fransen et al, 2003; Gratton and Vazquez, 2003; Ohlemiller, 2004; Gates and Mills, 2005). In addition, the genetic determinants responsible for late onset hearing loss are likely to be of only moderate effect; moderate inter-individual variations in gene expression or function at multiple loci masked, hindered or augmented by epistatic interactions and compounded by environmental interplay are thought to underlie susceptibility (for reviews see Jennings and Jones, 2001; Fransen et al, 2003). However, with careful study design case-control association analysis provides a powerful means to identify the genetic variants that underlie susceptibility to complex traits such as late onset hearing loss; importantly it provides great power to detect genes of moderate effect (Risch and Merikangas, 1996). Certainly, the case-control association study approach has been used successfully to delineate the genetic risk factors involved in other similar complex traits cardiovascular disease being a well-studied example (for reviews see Daley and Cargill, 2001; Green, 2001; Herrmann and Paul, 2002; Humphries et al, 2004; case-control association analysis as a method to decipher the genetic basis of complex disease is also discussed in section 1.6.2).

At the onset of this research project no studies had been published attempting to associate a candidate gene(s) with susceptibility to late onset hearing loss. During the course of this PhD project a few studies have been published that have failed to find an association between sequence variants in a candidate gene and susceptibility to late onset hearing loss (Van Laer et al, 2002; Fransen et al, 2004). Van Laer et al, 2002

focussed on DFNA5 as a potential candidate gene for late onset hearing loss. A mutation in DFNA5 causes progressive sensorineural hearing loss in a Dutch family which initiates in the high frequencies and is characteristic of late onset hearing loss, albeit manifesting at an earlier age of onset, 5-15 years (Van Laer et al, 1998). Two single nucleotide substitutions in the DFNA5 coding region leading to amino-acid substitutions were examined in two independent association based studies but no significant association to late onset hearing loss was found (Van Laer et al, 2002). However, in recent years some studies have been published in the literature reporting a positive association between sequence variants in a candidate gene and susceptibility to late onset hearing loss (Fortunato et al, 2004; Unal et al, 2005; Yang et al, 2006). Specifically, sequence variants in SOD2, PON2 and susceptibility to NIHL (Fortunato et al, 2004; see section 1.7.4), Cdh23 and susceptibility to NIHL (Yang et al, 2006; see section 1.7.1), and NAT2 and susceptibility to ARHL (Unal et al, 2005; see section 1.7.4) Although, in each of these studies a limited number of samples was used and the validity of these findings with respect to late onset hearing loss susceptibility in the general population requires replication in a larger cohort for confirmation.

The long-term goal of this research project is to use a case-control association study approach to investigate whether common sequence variants in the Brn-3c gene, an excellent candidate gene for late onset hearing loss are a risk factor for susceptibility to this disease (the evidence that Brn-3c is a good candidate gene for late onset hearing loss is discussed in section 1.8). Delineation of discrete genetic variants that contribute only a small amount to the overall phenotype of a complex disease such as late onset hearing loss is challenging. Unless the gene under study has a major affect on disease phenotype as is commonly observed with monogenic traits, a small sample size has limited power to detect a positive association (Risch and Merikangas, 1996). This is important, as it is unlikely that any one gene contributes to late onset hearing loss by a large extent given the paradigm of genetic heterogeneity observed for monogenic deafness (for reviews see Van Camp et al, 1997; Petersen, 2002) and the knowledge gained in analysis of other similar complex diseases (for review see Wright et al, 2003). However, use of moderately large and well-characterised sample cohorts in case-control association analysis of complex traits such as late onset hearing loss should in principle allow detection of even small genetic effects on phenotype (Risch and Merikangas, 1996; for reviews see Cardon and Bell, 2001; Hirschhorn et al, 2002 and Wright et al, 2003). Certainly, the merits of using moderately large sample sizes in case-control association studies has met with success in the analysis of other similar complex

diseases including cardiovascular disease (Cambien et al, 1992; for review see Herrmann and Paul, 2002), type-2 diabetes (Douglas et al, 2001) and schizophrenia (Shifman et al, 2002).

At the onset of this research project well-characterised patient cohorts for late onset hearing loss of any size were lacking. Hence, it was a major aim of the PhD project to initiate collection of a large, well-characterised patient cohort for late onset hearing loss that could be used in subsequent case-control association analysis (see section 1.10). It was accepted at the outset of the PhD project that the likelihood of finding a positive association would be limited by the power of the study, which in part would be dependent on the number of patients with late onset hearing loss it would be possible to recruit into this study within the time constraints of this project. Therefore, the case-control association analysis performed for this PhD project was considered a preliminary study as a first step prior to undertaking a large-scale population based case-control association study, which is beyond the scope of the PhD project, but an aim of longer term research.

8.2 Collection of a late onset hearing loss patient cohort.

At the onset of this project well-characterised patient cohorts for late onset hearing loss were scarce. Therefore, it was an important aim of this project to initiate collection of a large, well characterised patient cohort for late onset hearing loss that could be used in a preliminary association study as a first-step towards undertaking a large-scale population based case-control association study. Patients characterised by a consultant audiologist as having a late onset sensorineural hearing loss (for definition see section 1.5.1) were recruited from the adult hearing aid clinic at the Royal Free Nose, Ear and Throat Hospital, London, U.K (for more information see method section 3.2.1). Ethical approval for this part of the project was granted from the Royal Free Local Research Ethics Committee (6202).

Patients were recruited from the forthcoming clinic appointments lists with the following exclusions: less than 30 years of age, history of known dementia and / or learning disability or patient identified as requiring an interpreter (which are likely to cause problems with the consent procedure). Data sets collected on patients include gender, age, ethnicity, diagnostic audiogram, age of onset, family history of hearing loss, history of noise exposure including both acute (sudden) and chronic exposure (for example, occupational noise exposure) in addition to details of significant medical problems (see questionnaire in Appendix A). Upon review of patient data sets, patients were further excluded from the study if: they were non-Caucasians, upon diagnostic audiogram hearing loss was identified as conductive, a specific known aetiology of the sensorineural hearing loss was identified for example: acoustic neuroma (a non-cancerous growth near auditory nerves), Meniere's disease, evidence of head trauma leading to sudden sensorineural hearing loss or evidence of acute noise exposure leading to sudden sensorineural hearing loss. Patients were not excluded from the study if there was an asymmetric hearing loss as long as the better hearing ear was consistent with the criteria for late onset sensorineural hearing loss. In these cases onset of hearing loss was dated from the better hearing ear characteristic of late onset sensorineural hearing loss and the worst hearing ear (due to hearing loss manifesting from childhood infections, mastoid surgery or Meniere's disease) was ignored.

During the course of this PhD project 142 patients with late onset sensorineural hearing loss were recruited into this study that met the above inclusion criteria. Blood was collected in EDTA and genomic DNA extracted using standard phenol-chloroform extraction procedure (see method section 3.2.11.1).

The choice of the control population is a crucial factor in good study design for association-based analysis in complex disease (for review see, Cardon and Bell, 2001). Every effort was taken to use a control sample group as closely matched to the patient group as possible. However, recruitment of an aged- and ethnically- matched cohort of subjects with no medical history of hearing loss or any history of late onset hearing loss within their family is notoriously difficult to achieve. Use of parental family-based controls in association studies can be useful; one advantage of this approach is that it reduces the effect of population stratification resulting from ethnic admixture (for review, see Cardon and Bell, 2001). However, this approach is not appropriate given the late onset of the hearing loss phenotype and the likelihood of finding elder surviving family members. Therefore, for the preliminary case-control association study the control sample group consisted of genomic DNA samples that were obtained from random, healthy members of the general population and have been described previously (see Crawley et al, 1999). Data sets available on subjects include gender, age and ethnicity. For the preliminary case-control association study subjects of Caucasian ethnicity were selected; this gave 145 ethnically matched general population samples for use in association-based analysis. It was accepted at the onset of this project that a general population sample is not the 'ideal' control population but the 145 random, healthy members of the general population are the closest matching 'control' sample to the patient cohort that it was possible to obtain for the purpose of the preliminary association study.

8.2.1 Comparison of late onset hearing loss and general population cohorts.

The late onset hearing loss patient cohort initiated at the Royal Free Nose, Ear and Throat Hospital, London, U.K is extremely well characterised. However, limited information was available for the general population sample; using information that was available data sets were compared between the two sample groups.

The age range of the patients with late onset hearing loss included in the preliminary association study is from 39 to 90 years of age with an average age of 70-79 years exhibited for 37% of patients (see Table 8.1 and Fig. 8.1). This is much higher than that observed for subjects within the general population sample group: 17 to 58 years with an average age of 20-29 years exhibited for 61% of subjects (see Table 8.1 and Fig. 8.1). This difference in average age between patients with late onset hearing

Table 8.1: Comparison of age and gender between late onset hearing loss and general population cohorts.

	Late onset hearing loss	General population
Age range (years)	39 – 90	17 – 58
Average age (years)	70 – 79	20 – 29
Male (%)	49	44
Female (%)	50	50
Gender unknown (%)	1	6

All subjects were of Caucasian ethnicity. For more information on cohorts see figure legend 8.1

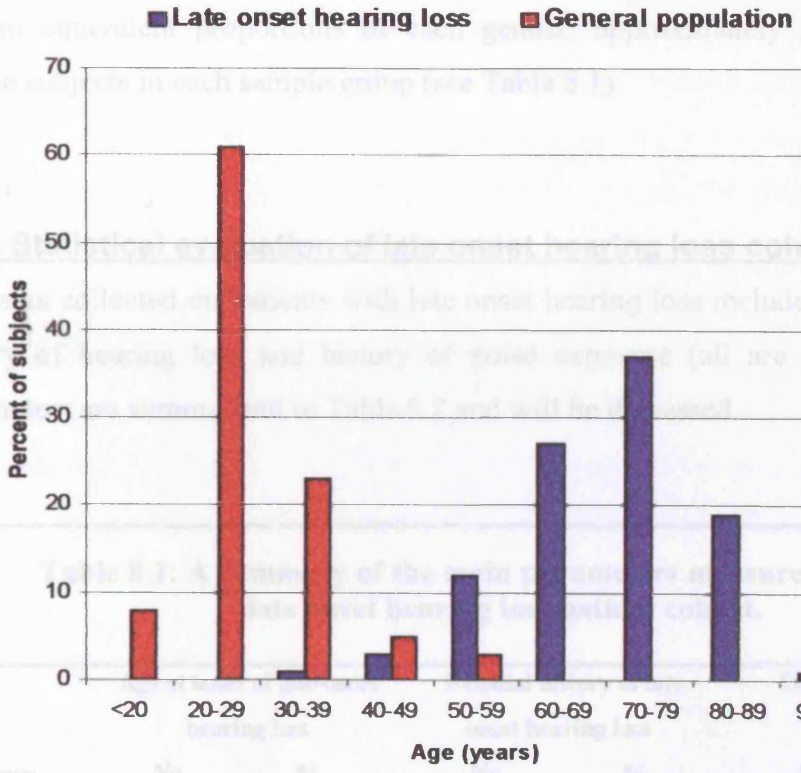


Figure 8.1 Comparison of age between late onset hearing loss and general population cohorts. Patients were recruited from the adult hearing aid clinic at the Royal Free Nose, Ear and Throat Hospital, London, U.K. All patients were characterised by a consultant audiologist as having a late onset hearing loss. General population samples were obtained from random, healthy members of the general population and have been described previously (see Crawley et al, 1999). All subjects were of Caucasian ethnicity.

loss and general population subjects is not ideal but an alternative control sample group was not available at the onset of this project. For the general population sample group any family history of late onset hearing loss was unknown and it was not possible to measure the hearing of subjects within this sample group. Hence, it was acknowledged that subjects in the general population sample group could be or may go on to be

affected by late onset hearing loss. To overcome this short-coming for future association studies recruitment of age-matched cohorts characterised with late onset hearing loss and good hearing is now underway in our laboratory (see section 8.8 for more information). However, it is important to realise that this discrepancy between the two sample groups for the preliminary association study is more likely to give a false-negative association rather than a false-positive association. Similarly, the likelihood of a false-positive association resulting from population stratification due to ethnic admixture is likely to be low in this preliminary association study as great care was taken to ensure that both patient and general population sample groups are ethnically matched. In addition, the likelihood of a false-positive association resulting from gender bias is unlikely to be encountered as both patient and general population sample groups contain equivalent proportions of each gender: approximately 50% male and 50% female subjects in each sample group (see Table 8.1).

8.2.2 Statistical evaluation of late onset hearing loss cohort.

Data sets collected on patients with late onset hearing loss include age of onset, family history of hearing loss and history of noise exposure (all are self-reported). These parameters are summarised in Table 8.2 and will be discussed.

Table 8.2: A summary of the main parameters measured within the late onset hearing loss patient cohort.

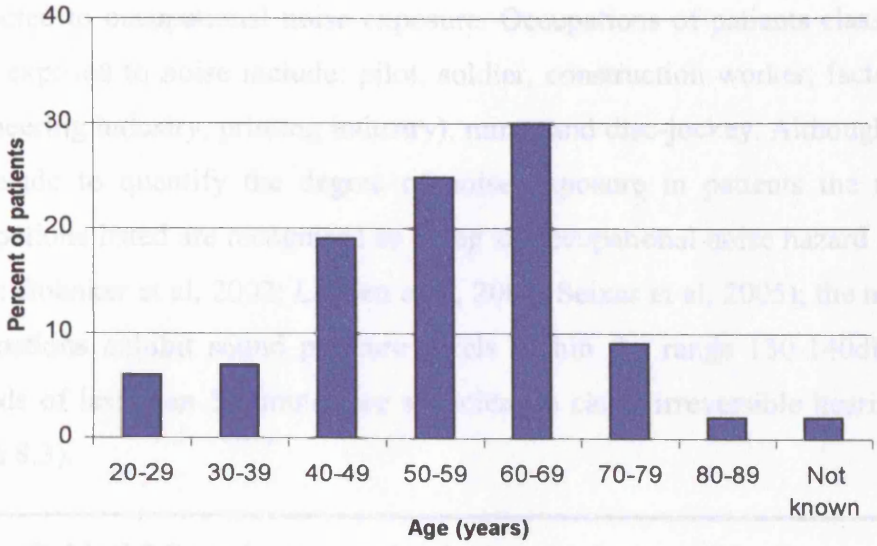
	Age at onset of late-onset hearing loss		Familial history of late onset hearing loss		Occupational noise exposure	
Years	**No.**	**%**	**No.**	**%**	**No.**	**%**
20-29	9	6 %	8	89 %	6	67 %
30-39	10	7 %	9	90 %	3	30 %
40-49	27	19 %	16	59 %	10	37 %
50-59	35	25 %	17	49 %	17	49 %
60-69	43	30 %	24	56 %	18	42 %
70-79	13	9 %	6	50 %	6	46 %
80-89	3	2 %	1	33 %	1	33 %
Unknown	2	2 %				
Total			**81 / 141***	**57 %**	**61 / 142**	**43 %**

* One patient within group 70-79 years did not know if there was a family history of late onset hearing loss.

Patients were categorised according to age at onset of late-onset hearing loss based on their own recollections of hearing loss (self-reported). The results of this analysis presented in Fig. 8.2a shows that in this patient sample the most common age for onset of late onset-hearing loss is 60-69 years of age, closely followed by 50-59 and 40-49 years of age. This is in agreement with the general concept that late onset hearing loss is increasingly common with advancing age (Davis, 1995). The heritability of late onset hearing loss is estimated to be 35-55% (Karlsson et al, 1997; Gates et al, 1999). Of the 142 patients with late onset hearing loss recruited into this study, 141 patients were able to provide information on their family history of hearing loss (self-reported). Of those patients where information was available, 57% have at least one close family member (defined as a grandparent, parent or sibling) that is also affected by a late onset hearing loss (see Table 8.2). When the percent of patients classified as having at least one close family member that is also affected by a late onset hearing loss is compared to the patients own age at onset of late-onset hearing loss it is clear that the younger the age of the patient when hearing loss first became apparent the greater the likelihood that the patient has at least one close family member that is also affected (Fig. 8.2b). For patients with hearing loss onset within the 20-29 and 30-39 year age groups 89% and 90%, respectively are classified as having at least one close family member that is also affected. In contrast, for patients with hearing loss onset within the elder age groups for example, the 70-79 and 80-89 year age groups only 46% and 33% respectively, are classified as having at least one close family member that is also affected (Fig. 8.2b). It is tempting to speculate that in the younger age groups heritable genetic factors have accelerated onset of late onset hearing loss.

Acoustic trauma is a major factor in development of late onset hearing loss (for review see Lynch and Kil, 2005). There is increasing evidence that ARHL and NIHL (hearing loss caused by prolonged exposure to moderate level sound) are not separate pathologies but overlapping phenomena (for reviews see Ohlemiller, 2004; Gates and Mills, 2005). It is thought that combinations of additive and non-additive effects explain the interplay between chronic noise exposure and ageing in the cochlea (for review see Rosenhall, 2003; for a more in-depth discussion see section 1.5.3, page 44). Acoustic trauma is widely recognised as a "global occupational health hazard" and worldwide, 16% of adult onset hearing loss classified as disabling is estimated to be due to occupational noise exposure (based on figures scored in 2000; see Nelson et al, 2005). Therefore, given the major role of acquired acoustic trauma in development of late onset hearing loss patients recruited into this study were surveyed for a history of noise

(a)



(b)

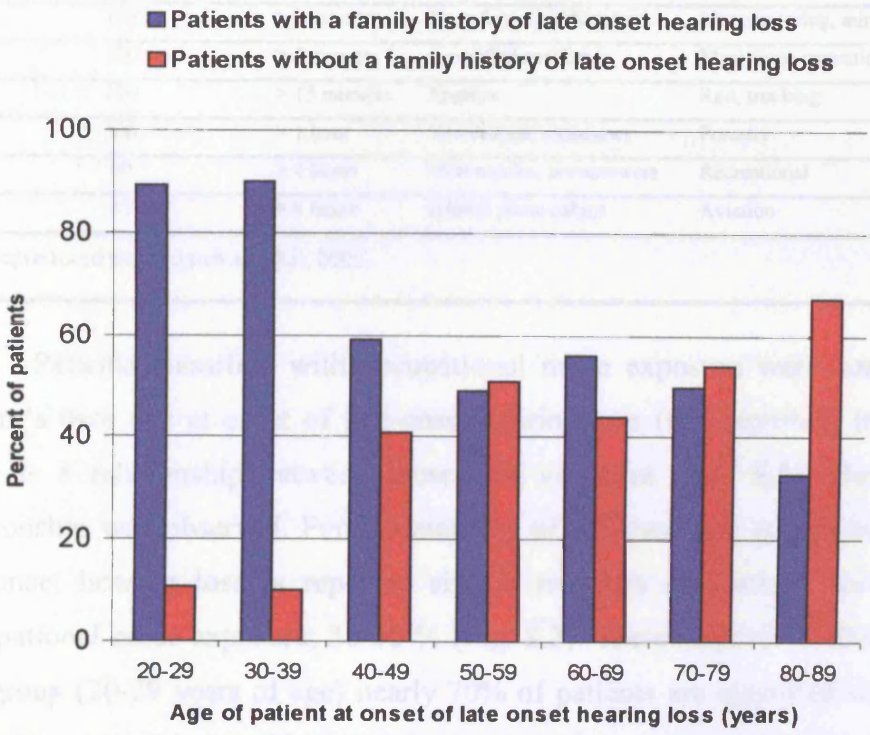


Figure 8.2: Effect of family history of late onset hearing loss on patients own age at onset of late onset hearing loss (self reported). (a) Age of patients at onset of late onset hearing loss. Patients were categorised according to age at onset of late-onset hearing loss based on their own recollections of hearing loss. **(b) Patient family history of late onset hearing loss.** Percent of patients that have at least one close family member with a late onset hearing loss compared to patients own age at onset of late-onset hearing loss. For comparison the percent of patients with no family history of late onset hearing loss is also shown.

exposure. For the purpose of this preliminary study attempts were not made to quantify the degree of noise exposure although the data set allows this to be done. Instead, patients were simply dichotomised (yes or no) according to whether they had been subjected to occupational noise exposure. Occupations of patients classified as having been exposed to noise include: pilot, soldier, construction worker, factory worker (eg: engineering industry, printing industry), miner and disc-jockey. Although attempts were not made to quantify the degree of noise exposure in patients the majority of the occupations listed are recognised as being an occupational noise hazard (Hong and Kim 2001; Bohnker et al, 2002; Landen et al, 2004; Seixas et al, 2005); the majority of these occupations exhibit sound pressure levels within the range 130-140dB and exposure periods of less than 5 minutes are sufficient to cause irreversible hearing damage (see Table 8.3).

Table 8.3 Sound pressure levels that can irreversibly damage hearing

Sound pressure level (dB)	Duration	Sound source	Industry
140	< 1 minute	Firearms, jet engine	Military, aviation
130	> 1 minute	Drop forge, jackhammers	Manufacturing, mining, construction
120	> 5 minutes	Amplified speaker	Musicians, recreational
110	> 15 minutes	Engines	Rail, trucking
100	> 1 hour	Woodshops, chainsaws	Forestry
90	> 4 hours	Motorcycles, lawnmowers	Recreational
85	> 8 hours	Interior plane cabins	Aviation

Reproduced from Lynch and Kil, 2005.

Patients classified with occupational noise exposure were compared with the patient's own age at onset of late-onset hearing loss (self-reported) to assess whether there is a relationship between these two variables (Fig. 8.3). However, no clear relationship was observed. For the majority of different age groups at which onset of late-onset hearing loss is reported similar numbers of patients are classified with occupational noise exposure; 30-50 % (Fig. 8.3). The exception to this is the youngest age group (20-29 years of age) nearly 70% of patients are classified with occupational noise exposure; it is possible that noise exposure has accelerated onset of hearing loss in this sample group. It would be interesting to extend this line of analysis. Recruitment of subjects into our late onset hearing loss cohort from the adult hearing aid clinic at the Royal Free Nose, Ear and Throat Hospital, London, U.K. has now accumulated to 500 patients and a detailed history of noise exposure is available for each patient. It would be interesting to explore the relationship between age at onset of hearing loss and degree

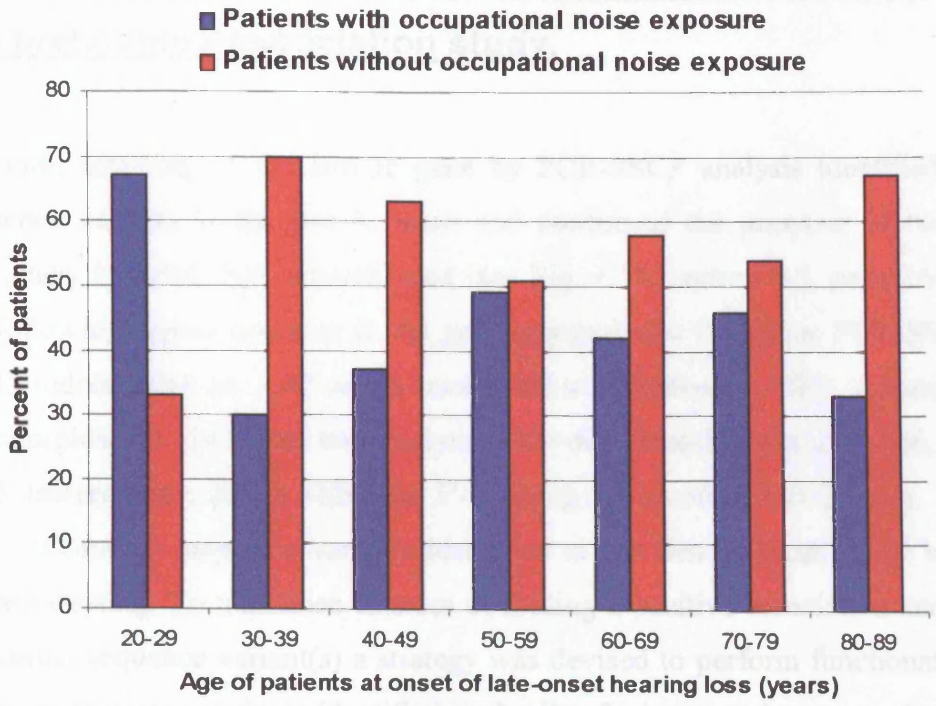


Figure 8.3 Percent of patients classified with occupational noise exposure compared to patients own age at onset of late-onset hearing loss (self-reported). For comparison percent of patients with no history of occupational noise exposure is also shown.

of noise exposure in an attempt to delineate the effect of noise exposure on age at onset of hearing loss.

In summary, 142 patients characterised by a consultant audiologist as having a late onset sensorineural hearing loss were recruited into the preliminary association study within the time constraints of this project. Recruitment of a suitable control sample group at the onset of this project was difficult and consisted of 145 random, healthy members of the general population. Recruitment of age-matched cohorts characterised with late onset hearing loss and good hearing is now underway in our laboratory and will be a valuable resource for future association studies. All subjects used in the preliminary association study are of Caucasian ethnicity.

8.3 Common sequence variants at the Brn-3c locus selected for the preliminary association study.

Mutation scanning of the Brn-3c gene by PCR-SSCP analysis identified five novel sequence variants at the Brn-3c locus and confirmed the presence of two that were previously reported, but non-validated (see Fig. 4.16, section 4.5, page 162). Of these the following appear common in the general population (based on PCR-SSCP analysis on 45 individuals): an A>C single nucleotide substitution at -1391, a complex poly-G polymorphism at -3432 and three polymorphic dinucleotide repeats at -566, -3457 and -3495. Interestingly, all are within the 5'-flanking region of the Brn-3c gene.

Common sequence variants identified at the Brn-3c locus could be neutral or disease causing. To maximise chances of finding a positive association and identifying the casual sequence variant(s) a strategy was devised to perform functional analysis on common sequence variants identified at the Brn-3c locus and to use only those where there is evidence to implicate they are functional in subsequent case-control association analysis (see section 1.10). By using this *functional screening approach* a false-positive association arising from identification of a neutral sequence variant that is in linkage disequilibrium with the casual variant at the same or another locus is minimised. In addition, this approach also allows for more reliable interpretation of case-control association data as it is only when a casual sequence variant(s) has been functionally characterised that its pathogenic role can be fully assessed and its effect on disease risk can be elevated with confidence. Problems associated with multiple testing and finding an association by chance (type I error) are also minimised.

For functional screening a range of *in-vitro* analysis was performed utilising EMSA analysis to assess whether common sequence variants in the Brn-3c 5'-flanking region modify the binding of transcription factors and transient transfection reporter gene assay to assess whether differences in binding affinity translate into functional differences in Brn-3c promoter activity (see Chapters 5 to 7). The results of this analysis suggests that the following sequence variants could have a functional effect: -3432poly-G polymorphism, -1391A >C and the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5' haplotype.

The -3432poly-G polymorphism modifies the binding affinity of an OC-2 derived nuclear protein (protein A) and there is convincing evidence that this is the transcription factor SP1 or at least a member of this transcription factor family. Use of

purified human recombinant SP1 protein, *in-vitro* translated SP1 and *in-vitro* translated SP3 confirms that the -3432polyG polymorphism modulates a high affinity SP family binding site and evidence suggests that this alters the regulation of the Brn-3c promoter when SP1 levels are limiting, $P<0.05$ (paired t-test; see Fig. 5.13b, page 203). Moreover, the data suggest a functional interaction between the -3432poly-G polymorphism and the -566$(GT)_n$ repeat which associate to determine the response of the Brn-3c gene to SP1. Similarly, the -1391A >C single nucleotide substitution modifies the binding affinity of an OC-2 derived nuclear protein, protein J. The variant allele, C has a reduced affinity for nuclear protein J compared to the wild-type allele, A and this appears to have an affect, albeit subtle on basal activity of the Brn-3c promoter. Basal activity from the Brn-3c promoter-luciferase reporter gene construct that carries the variant C allele at position -1391 is 26% less compared to that which carries the wild-type A allele and this difference although subtle is statistically significant, $P<0.05$ (t-test; Fig. 7.8b, page 257). Therefore, for the purpose of this PhD project it was decided to genotype both the -3432poly-G polymorphism and the -1391A>C single nucleotide substitution for the preliminary association study to establish allele frequency and to assess if either (or both) of these common sequence variants at the Brn-3c locus for which there is evidence to suggest they are functional, are a risk factor for late onset hearing loss as far as could be determined within the power of the preliminary study.

8.4 Genotyping the -3432poly-G polymorphism.

8.4.1 Genotyping strategy.

The -3432poly-G polymorphism in the Brn-3c promoter is an extremely complex polymorphism exhibiting multiple variations in length coupled with single nucleotide substitutions within the poly-G repeat (see Table 4.3, section 4.4.2, page 156). A search of the literature revealed that the -3432polyG polymorphism is very similar to a complex poly-G length polymorphism that has been identified in the promoter of the KLK1 gene which has a functional affect on activity of the KLK1 promoter and is associated with hypertensive end stage renal disease (Song et al, 1997; Yu et al, 2002; see also section 5.2).

To genotype the poly-G length polymorphism in the KLK1 gene promoter Song et al, and Yu et al, used direct sequencing for primary analysis. However, distinct alleles were sometimes difficult to define in heterozygous individuals due to the complexity of the KLK1 poly-G polymorphism (multiple variations in length of the poly-G repeat in addition to multiple single nucleotide substitutions within the poly-G repeat). To overcome the difficulties presented when genotyping complex heterozygotes a novel approach was devised to manually interpret the sequencing data (Yu et al, 2002). Discrete alleles were elucidated by first locating an invariant sequence downstream of the KLK1 poly-G polymorphism and reading the sequencing chromatogram manually from this point in the direction 5' to 3'; once the sequence downstream of the KLK1 poly-G polymorphism had been clarified it was possible to manually interpret each poly-G allele in the direction 3' to 5'.

From examination of the promoter sequence flanking the -3432poly-G polymorphism (Fig. 8.4) it is evident that any technique chosen to genotype the -3432poly-G polymorphism will be limited by the proximity of the multi-allelic dinucleotide repeats at positions -3457 and -3495: $(GA)_{1\text{-}3}$ and $(GT)_{15\text{-}21,\,24,}$ respectively. Due to the success of Song et al, and Yu et al, with direct sequencing in their analysis of the poly-G length polymorphism in the KLK1 promoter it was decided to use direct sequencing to genotype the -3432poly-G polymorphism in the Brn-3c promoter and to adopt the method of Yu et al, to manually interpret the sequencing data in heterozygous subjects.

8.4.1.1 Primer design and optimisation of genotyping.

The design of primers for use in genotyping the -3432poly-G polymorphism by direct sequencing was complicated by the nature of the Brn-3c promoter in the vicinity of the -3432poly-G polymorphism. The region of the Brn-3c promoter surrounding the -3432poly-G polymorphism is highly repetitive; in addition to the -3457$(GA)_n$ and -3495$(GT)_n$ dinucleotide repeats many short mononucleotide repeats are evident (Fig. 8.4). This highly repetitive nature of the DNA sequence severely restricted the choice and location of possible primers to use in genotyping the -3432poly-G polymorphism by direct sequencing.

Given the complexity of the Brn-3c promoter in the vicinity of the -3432poly-G polymorphism a protocol for genotyping the -3432poly-G polymorphism by direct sequencing was devised and optimised using predominantly primers that were already available in this region. Using the primers originally devised for SSCP analysis of the -3495$(GT)_n$ repeat (Fig. 8.4; highlighted in blue), for investigation of the -3432poly-G polymorphism in combination with the -3495$(GT)_n$ and -3457$(GA)_n$ repeats (Fig. 8.4 highlighted in green) and a novel primer, G3cP13-AS (Fig 8.4; highlighted in purple on the anti-sense strand) several different primer combinations were tested to PCR up an

```
GGAGGATATGGTGCTTTGTGCCCTGACTTCACCTTACATGTGGCAAATCTATGTCCCTACTAGACTCTGA -3742 to
CCTCCTATACCACGAAACACGGGACTGAAGTGGAATGTACACCGTTTAGATACAGGGATGATCTGAGACT -3673

GCTCCTGAAGCAGTGTCTTCTGTATCTTCAAGGCCTTCCACAAGACTCTGCTGTTAGTAGGTATTTGGAA -3672 to
CGAGGACTTCGTCACAGAAGACATAGAAGTTCCGGAAGGTGTTCTGAGACGACAATCATCCATAAACCTT -3603

ATGCTTATCGAATGGGGAAAATTCTTCCTTCAATGCCACTTGCATAAATGCCCACATAGTCCTTACATGG -3602 to
TACGAATAGCTTACCCCTTTTAAGAAGGAAGTTACGGTGAACGTATTTACGGGTGTATCAGGAATGTACC -3533

ACCTTCACATATTTTGTTTGGCCCATAAATGGTATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG -3532 to
TGGAAGTGTATAAAACAAACCGGGTATTTACCATACTCACACACACACACACACACACACACACACAC -3463

TGTGTGATTGTAATTTAATGCCATGGTGGTGGGGGGGGGGGTGGGTAAGTGCTCTCTAGTTTACCACAGT -3462 to
ACACACTAACATTAAATTACGGTACCACCACCCCCCCCCCACCCATTCACGAGAGATCAAATGGTGTCA -3393

TTCAATCTTTCCTTTTATCCACCCAGATAAAGTTGCTCAATTTTATTTTACTAGTTTTGCACCTTATAGA -3392 to
AAGTTAGAAAGGAAAATAGGTGGGTCTATTTCAACGAGTTAAAATAAAATGATCAAAACGTGGAATATCT -3323

ATTTAAGTTTACAAATTCTGGAAAGCATAGAGGCTGTTACAAATTGCCGTCTAAGGCTGTCCTTCCAATT -3322 to
TAAATTCAAATGTTTAAGACCTTTCGTATCTCCGACAATGTTTAACGGCAGATTCCGACAGGAAGGTTAA -3253

GCTTTTTCTATGGTCTTTTCCCTCTTTTCTTCTCCAGTCTCCAGTCTGTTATCTCCAAGGCCTTCTCAAA -3252 to
CGAAAAAGATACCAGAAAAGGGAGAAAAGAAGAGGTCAGAGGTCAGACAATAGAGGTTCCGGAAGAGTTT -3183
```

Figure 8.4 Sequence of the Brn-3c promoter in the vicinity of the -3432poly-G polymorphism to show possible primer combinations for genotyping. The -3432poly-G polymorphism and the -3495$(GT)_n$ repeat are highlighted in black bold; the -3457$(GA)_n$ repeat in red bold. The region of the Brn-3c promoter in the vicinity of the -3432poly-G polymorphism is highly repetitive. In addition to the -3457 and -3495 dinucleotide repeats many short mononucleotide repeats are evident. This repetitive nature of the Brn-3c promoter hinders the location of possible primers. Sequence of primers already available in this region are: 3cP11-S and 3cP11-AS used for PCR-SSCP analysis of the -3495$(GT)_n$ repeat (highlighted in blue with direction of each primer defined by the blue arrows; see also section 4.2.2) and 3cP12-S and 3cP12-AS devised to investigate the -3432poly-G polymorphism in combination with the -3495$(GT)_n$ repeat and -3457$(GA)_n$ repeat (highlighted in green with direction of each primer defined by the green arrows; see also section 4.2.2). Sequence of novel primer G3cP13-AS is highlighted in purple with direction of primer defined by the purple arrow.

amplicon containing the -3432poly-G polymorphism (see Table 8.4).

For each PCR reaction with a different primer combination genomic DNA of known -3432 genotype was used and the following parameters were varied: the magnesium concentration, the primer concentration and the annealing temperature of the reaction. Sub-optimal results were obtained using *Taq* DNA polymerase (Promega) even when the PCR reaction was supplemented with 5-10% DMSO; the intensity of the PCR bands ranged from faint to moderate, or were not evident. However, excellent and consistent results were obtained when the DNA polymerase was changed to AmpliTaq Gold (Applied Biosystems); an intense PCR band was observed for all primer combinations tested.

Having achieved specific and reproducible PCR amplification of genomic DNA for the region spanning the -3432poly-G polymorphism with several different primer sets the PCR products were purified with QiaQuick™ purification columns (Qiagen) and different primers (either 3cP12-AS, G3cP13-AS or 3cP12-S as outlined in Table 8.4) were tested to sequence the -3432poly-G polymorphism on an ABI 3730 DNA analyser using BigDye terminator chemistries version 3.1. The results of this analysis revealed that the primer combination 3cP11-S and 3cP12-AS in the PCR followed with the anti-sense primer, 3cP12-AS, in sequencing produces clear and consistent sequencing results for the -3432poly-G polymorphism; good quality sequence data with a strong signal is obtained 18bp from the 3' end of primer 3cP12-AS. Other primer combinations that were tested were not as successful. For example, the primer combination 3cP12-S and 3cP12-AS (or G3cP13-AS) in the PCR followed with the anti-sense primer 3cP12-AS for sequencing produces sequencing data that is confounded by a high background. In addition, it became apparent from this analysis that due to the proximity of the -3457$(GA)_n$ and -3495$(GT)_n$ repeats it is not possible to sequence the -3432poly-G polymorphism using a sense primer; the quality of the

Table 8.4: Primer combinations considered for genotyping the -3432poly-G polymorphism by direct sequencing.

Sense primer	Anti-sense primer	Size of PCR amplicon (bp)	Sequencing primer
3cP12-S	3cP12-AS	398	3cP12-AS
3cP12-S	G3cP13-AS	423	3cP12-AS
3cP12-S	G3cP13-AS	423	3cP12-S
3cP12-S	G3cP13-AS	423	G3cP13-AS
3cP11-S	3cP12-AS	285	3cP12-AS
3cP11-S	G3cP13-AS	310	3cP12-AS

sequence data 3' of the -3495$(GT)_n$ and -3457$(GA)_n$ repeats and prior to the -3432poly-G polymorphism was consistently poor when this approach was tried. Once a successful primer combination with which to genotype the -3432poly-G polymorphism by direct sequencing was established final adjustments were made to optimise this analysis. The amount of purified PCR product per sequencing reaction was adjusted for optimum peak height of the sequencing chromatograms and 100ng of purified DNA was found to give optimum results.

In summary, by careful selection of PCR and sequencing primers a suitable protocol (described in detail in method section 3.2.15.1) was devised and optimised to genotype the -3432poly-G polymorphism by direct sequencing and problems in sequencing that derived from the close proximity of the -3457$(GA)_n$ and -3495$(GT)_n$ repeats were minimised.

8.4.1.2 Manual interpretation of sequencing chromatograms.

To determine the -3432poly-G genotype in heterozygous subjects the method of Yu et al, was adopted and sequencing chromatograms were interpreted manually. The methodology for manual interpretation of sequencing chromatograms is described below and for clarity the majority of this text is reproduced in figure legend 8.5.

Manual interpretation of sequencing chromatograms in heterozygous subjects was aided by first locating the invariant sequence 'TGGCATTAAATTA' immediately downstream of the -3432 poly-G-polymorphism (Fig. 8.5; underlined for each allele within the written sequence added below the chromatogram). By identifying this invariant sequence first and reading the chromatograms 3' to 5' across the poly-G-polymorphic region (shown as a series of constitutive C's in Fig.8.5 as sequence obtained is for the anti-sense strand) it was possible to categorise homo- and heterozygous peaks and clarify each allele. Specifically, the invariant sequence was located by identifying the initial three bases 'TGG' which in heterozygous individuals is easily identified as a region of either 3, 4 or 5 constitutive G's (shown as 3 constitutive G's in the example shown in Fig. 8.5). Then counting from the 3' G (Fig. 8.5; highlighted in blue for allele 1 within the written sequence added below the chromatogram; also highlighted in blue for allele 1 within the text) the first (and longest) allele sequence was marked off the chromatogram in the direction 3' to 5' using the reference sequence (Ensembl transcript ID: ENST00000230732; see Appendix B) and poly-G allele sequences determined from the PCR-SSCP screen as a guide (see Table 4.3, section 4.4.2, page 156). Referring back to the 3' G of the

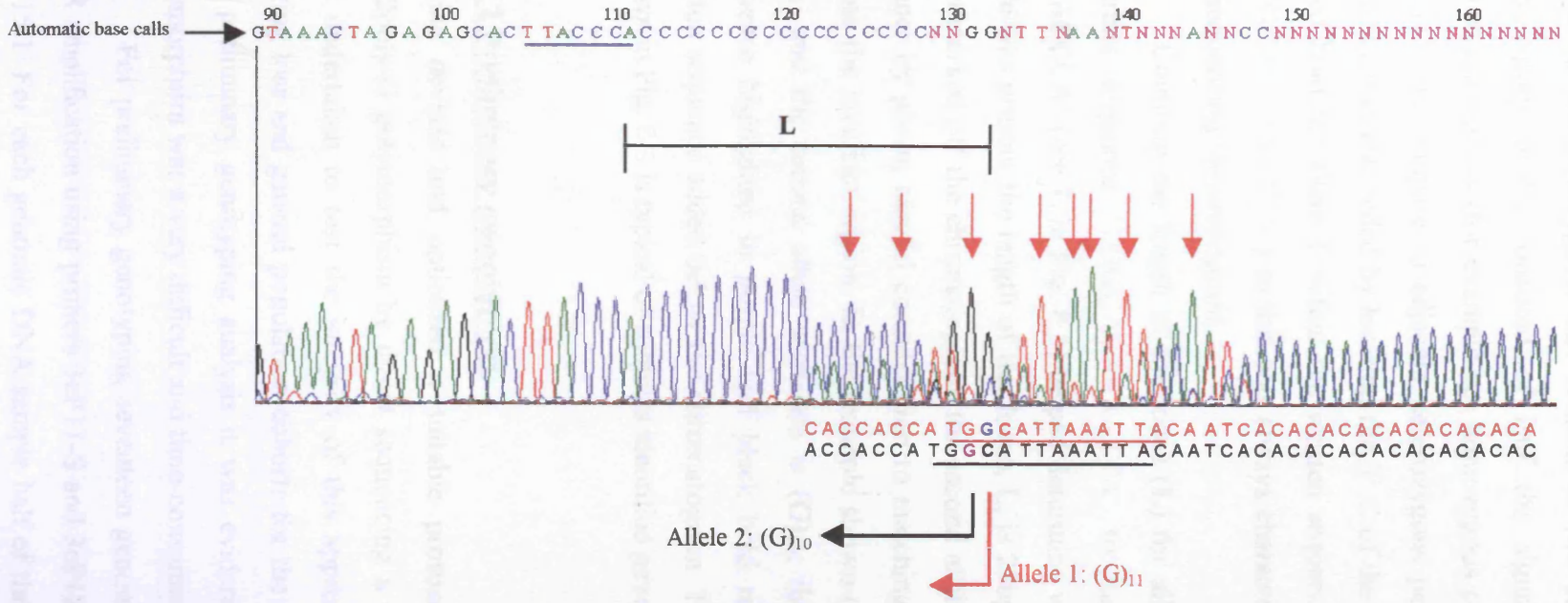


Figure 8.5: Representative sequencing chromatogram to illustrate the methodology used to manually interpret discrete -3432poly-G alleles. The -3432poly-G polymorphism was genotyped by direct sequencing with anti-sense primer 3cP12-AS and sequence obtained was for the anti-sense strand. Chromatograms obtained from direct sequencing were manually interpreted in heterozygous subjects to elucidate discrete alleles. Manual interpretation of the sequencing chromatogram was aided by first locating the invariant sequence 'TGGCATTAAATTA' immediately downstream of the -3432poly-G polymorphism (underlined in red for allele 1 and black for allele 2 within the written sequence added below the chromatogram). Specifically, the invariant sequence was located by identifying the initial three bases 'TGG' which in heterozygous individuals consists of a region of either 3, 4 or 5 constitutive G's (shown as 3 constitutive G's in above example). Then counting from the 3' G (highlighted in blue for allele 1 within the written sequence added below the chromatogram) the first (and longest) allele sequence was marked off the chromatogram in the direction 3' to 5'. Referring back to the 3' **G** of the invariant sequence for allele 1 the rest of the invariant sequence 'CATTAAATTA' was marked off the chromatogram in the direction 5' to 3'. Identification of this remaining invariant region was aided by clear distinction between homozygous and heterozygous peaks within this region (homozygous peaks that facilitated identification of each allele in the above example are indicated by the small red arrows). Identification of the invariant sequence was also aided by locating the 3' C of the invariant sequence (shown in red bold for allele 1 within the written sequence added below the chromatogram) as this was characterised by a very low peak height on the sequencing chromatogram. Counting the length of the locus (**L**) for allele 1 defined from the 3' **G** of the invariant sequence 'TG**G**CATTAAATTA' to the 3'**A** of the upstream sequence 'TTACCC**A**' (underlined in blue for the automatic base calls) helped determine whether allele 1 was $(G)_{12}$; if the $(G)_{12}$ allele is present the length of the locus, L, is 23bp. In the example shown L = 22bp and allele one (the longest allele) is $(G)_{11}$. Once the first allele sequence had been marked off the chromatogram the second allele sequence was deduced in a similar manner by giving careful consideration to matching of homo- and heterozygous peaks across the invariant region. In the example shown for allele 2 the 3' **G** of the invariant sequence is highlighted in purple and the 3' **C** of the invariant sequence is highlighted in black bold within the written sequence added below the chromatogram. The direct sequencing chromatogram shown is typical of the subjects assigned genotype $(G)_{11}$ / $(G)_{10}$.

invariant sequence for allele 1 the rest of the invariant sequence 'CATTAAATTA' was marked off the chromatogram in the direction 5' to 3'. Identification of this remaining invariant sequence was aided by the clear distinction between homozygous and heterozygous peaks within this region; homozygotes are characterised by a single peak compared to heterozygotes that consists of two superimposed peaks. In addition the heterozygous peaks consisted of half the signal strength compared to adjacent homozygous peaks (for example see homozygous peaks indicated by the red arrows in Fig. 8.5 and compare to adjacent heterozygous peaks). Identification of the invariant sequence was also aided by locating the 3' C of the invariant sequence (Fig. 8.5; shown in red bold for allele 1 within the written sequence added below the chromatogram; 'TGGCATTAAATTA') as this was always characterised by a very low peak height on the sequencing chromatogram.

Counting the length of the locus (**L**) for allele 1 defined from the 3' **G** of the invariant sequence 'TGGCATTAAATTA' to the 3'**A** of the upstream sequence 'TTACCCA' (see **L** in Fig. 8.5) helped determine whether allele 1 is $(G)_{12}$. If the $(G)_{12}$ allele was present the length of this locus, L, is 23bp. Once the first allele sequence had been marked off the chromatogram, the second allele sequence was deduced in a similar manner by giving careful consideration to matching of homo- and heterozygous peaks across the invariant region. In the example shown (Fig. 8.5) the first allele sequence is $(G)_{11}$ and the second allele sequence is $(G)_{10}$; the 3' G and 3' C of the invariant sequence highlighted in purple and black bold respectively, for allele 2 within the written sequence added below the chromatogram. The direct sequencing chromatogram shown in Fig. 8.5 is typical of subjects identified genotype $(G)_{11}$ / $(G)_{10}$.

8.4.2 Preliminary genotyping.

Having devised and optimised a suitable protocol with which to genotype the -3432poly-G polymorphism by direct sequencing a small set of preliminary genotyping was undertaken to test the integrity of this approach prior to genotyping late onset hearing loss and general population cohorts for the preliminary association study. From this preliminary genotyping analysis it was evident that genotyping the -3432poly-G polymorphism was a very difficult and time-consuming process.

For preliminary genotyping seventeen genomic DNA samples were subjected to PCR amplification using primers 3cP11-S and 3cP12-AS as described in method section 3.2.15.1. For each genomic DNA sample half of the PCR product was genotyped for -

-3432poly-G polymorphism with anti-sense primer 3cP12-AS according to method section 3.2.15.1 and half was sub-cloned into a T-Tailed vector, pGEM®-T easy (Promega) (see method section 3.2.9.1) to allow sequence analysis of individual DNA molecules from the PCR product pool. Several clones for each sample were sequenced for -3432poly-G polymorphism using primer 3cP12-AS and results obtained were compared with those obtained from direct sequencing.

Using direct sequencing coupled with manual interpretation of sequencing chromatograms it was evident that discrete poly-G alleles were sometimes difficult to elucidate in heterozygous individuals with different length variants. However, it was clear that heterozygous samples could be grouped according to genotype based on reproducible and consistently similar sequencing chromatograms. This problem encountered in genotyping complex heterozygotes for the -3432poly-G polymorphism by direct sequencing is not unique; it is similar to that described by Yu et al, in their analysis of the poly-G length polymorphism in the KLK1 promoter.

Sequencing of cloned PCR products did not aid elucidation of discrete poly-G alleles in complex heterozygotes. Similarly, neither did it prove to be a valuable method to confirm primary genotyping data obtained by direct sequencing and this was especially evident if a subject was heterozygous for the -3432poly-G polymorphism; poly-G alleles elucidated from sequencing of cloned PCR products did not always clearly agree with data obtained by direct sequencing. 'Stutter' alleles from the PCR had a tendency to be sub-cloned alongside the bona-fide genuine alleles making discrimination of genuine alleles from PCR artefacts difficult. Although, it was evident that if a subject was homozygous for the -3432poly-G polymorphism sequencing of cloned PCR products was often consistent with data obtained by direct sequencing particularly if the shortest poly-G alleles were present. A detailed discussion of this line of analysis is presented in Appendix G.

Given the overall unsuitability of sequencing cloned PCR products to confirm and / or to aid identification of -3432poly-G alleles determined by direct sequencing an alternative confirmatory technique was sought. In some heterozygous samples identified as having the same genotype based on reproducible and consistently similar sequencing chromatograms it was evident that if the size of the two poly-G alleles could be ascertained, then elucidation of the precise sequence for each allele would be facilitated. Therefore, to aid genotyping in these difficult sample groups and as a confirmatory method allele size was determined in each of the original 17 genomic DNA samples by

polyacrylamide gel electrophoresis (PAGE) analysis of radioactive PCR products. This line of analysis is outlined in the next section.

8.4.2.1 PAGE analysis of radioactive PCR products as a confirmatory method for genotyping the -3432poly-G polymorphism by direct sequencing.

To facilitate identification of discrete -3432poly-G alleles in complex heterozygotes and as a confirmatory method for genotyping the -3432poly-G polymorphism by direct sequencing PAGE analysis of radioactive PCR products was used to size poly-G alleles in each of the original 17 genomic DNA samples selected for preliminary genotyping. This approach proved to be very valuable to aid elucidation of discrete poly-G alleles and to confirm genotypes identified by direct sequencing.

To size poly-G alleles by PAGE analysis of radioactive PCR products a novel primer G3cP13-S: 5'-GATTGTAATTTAATGCCATGGTG-3' was designed for use in combination with primer 3cP12-AS to amplify a 181bp* product containing the -3432poly-G polymorphism (*based on containing a -3432poly-G allele of $(G)_{11}$; see Appendix B for reference sequence). By careful placement of the sense primer, G3cP13-S, it was possible to generate PCR amplicons of length dependent on the nature of the -3432poly-G allele and not the proximal -3457$(GA)_n$ or -3495$(GT)_n$ polymorphic dinucleotide repeats (see Fig. 8.6). A summary of the different size -3432poly-G alleles generated is shown in Table 8.6 (sizes are shown for all nine alleles identified at the -3432 locus) and full details of the protocol is described in method section 3.2.15.2.

In summary, preliminary genotyping analysis suggested that direct sequencing is a reliable method with which to genotype the -3432poly-G polymorphism if sizing of poly-G alleles by PAGE analysis of radioactive PCR products is used as an additional secondary technique to aid elucidation and / or to confirm discrete alleles in complex heterozygotes. Sequencing of cloned PCR products as a confirmatory method did not prove valuable. Based on these results it was decided to genotype the -3432poly-G polymorphism in late onset hearing loss and general population cohorts using direct sequencing as the primary analysis and then where necessarily to perform secondary analysis to size poly-G alleles to confirm or where appropriate facilitate identification of genotypes.

Figure 8.6 Section of the Brn-3c promoter surrounding the -3432 poly-G polymorphism to show sequence of primers for sizing poly-G alleles by PAGE analysis of radioactive PCR products. The region of the Brn-3c promoter in the vicinity of the -3432poly-G polymorphism is highly repetitive consisting of two multi-allelic dinucleotide repeats: -3457$(GA)_n$ and -3495$(GT)_n$ highlighted in red bold and black bold, respectively (the -3432poly-G polymorphism is also highlighted in black bold). By careful primer design it was possible to size poly-G alleles by PAGE analysis of radioactive PCR products and exclude the effect of the polymorphic dinucleotide repeats at -3457 and -3495 from the analysis. Sequences of primers used are highlighted in green (3cP12-AS) and purple (G3cP13-S) with the direction of each primer defined by the arrows. The G3cP13-S primer includes the 3' two nucleotides of the -3457$(GA)_n$ repeat.

Table 8.5: A summary of the different sized -3432poly-G alleles generated by PCR amplification of genomic DNA.

Allele	DNA sequence	Abbreviation	Length of allele (bp)*
$(G)_{10}$	5' TGGTGGT**GGGGGGGGGG**TGGGTA 3'	$(G)_{10}$	181
$(G)_{11}$	5' TGGTGGT**GGGGGGGGGGG**TGGGTA 3'	$(G)_{11}$	182
$(G)_{12}$	5' TGGTGGT**GGGGGGGGGGGG**TGGGTA 3'	$(G)_{12}$	183
$(G)_{13}$	5' TGGTGGT**GGGGGGGGGGGGG**TGGGTA 3'	$(G)_{13}$	184
SNPG1	5' TGGTGGT**GGGGGGGGC**GTGGGTA 3'	$(G)_8CG$	181
SNPG2	5' TGGTGGT**GGC**GGGGGGGGGTGGGTA 3'	$GGC(G)_9$	183
SNPG4	5' TGGTGGT**GGT**GGGGGGGGTGGGTA 3'	$GGT(G)_8$	182
SNPG5	5' TGGTGGT**GGGGC**GGGGGGTGGGTA 3'	$(G)_4C(G)_6$	182
SNPG6	5' TGGTGGT**GGC**GGGGGGGGGGTGGGTA 3'	$GGC(G)_{10}$	184

* Allele length determined from using primers G3cP13-S and 3cP12-AS in the PCR reaction.

8.4.3 Genotyping the -3432poly-G polymorphism in late onset hearing loss and general population cohorts.

Having devised and optimised a suitable protocol for genotyping the -3432poly-G polymorphism by direct sequencing and having tested the integrity of this method and made adjustments the -3432poly-G polymorphism was genotyped in 142 patients characterised as having a late onset sensorineural hearing loss and 145 general population samples for a preliminary association study.

Direct sequencing was used as the primary analysis. Upon direct sequencing samples were initially grouped according to chromatogram data and in the case of complex heterozygotes, similar chromatogram patterns. PAGE analysis of radioactive PCR products was performed in secondary analysis to size poly-G alleles to facilitate identification of discrete alleles in complex heterozygotes and to confirm the genotypes determined from primary analysis. In all samples where poly-G alleles were sized the results of sizing poly-G alleles by PAGE analysis of radioactive PCR products aided elucidation of genotype by direct sequencing with a few exceptions. Firstly, for subjects assigned to chromatogram group 3 and identified as $(G)_{12}$ heterozygotes (see Table 8.6) it was not possible to discriminate between $(G)_{11}$ and $(G)_{10}$ as the second allele type and so in frequency analysis both possibilities were counted. Secondly, it was not possible to clarify the second allele type for subjects with genotype SNPG4 / ND (**N**ot **D**etermined; see Table 8.6) and therefore, this category of subjects was excluded from subsequent frequency analysis. In both these cases the presence of convincing 'shadow' bands upon PAGE analysis of radioactive PCR products confounded accurate sizing. Similarly, for a small number of samples (3 samples) identified SNPG2 homozygous by direct sequencing sizing of poly-G alleles by PAGE analysis of radioactive PCR products did not agree with the data obtained by direct sequencing. The presence of a convincing secondary band upon PAGE analysis of radioactive PCR products prevented clarification of genotype. Consequently the genotype could not be identified in these samples and they were left unassigned (see Table 8.6) and excluded from subsequent frequency analysis. A detailed discussion of these problems encountered in genotyping the -3432poly-G polymorphism is discussed further in Appendix H: *elucidation of problematic genotypes.*

For some samples the quality of the data obtained from direct sequencing was poor; either a high background or loss of sequencing signal after the poly-G repeat confounded reliability of the data or made accurate determination of discrete poly-G alleles impossible. These samples were immediately subjected to re-sequencing and if

Table 8.6: A summary of -3432poly-G genotypes identified in late onset hearing loss and general population cohorts using direct sequencing coupled with secondary analysis.

Chromatogram group	Genotype	Late onset hearing loss (observed)	General population (observed)
1	$(G)_{11}$ / $(G)_{10}$	69	52
2	$(G)_{10}$ / $(G)_{10}$	3	4
3	$(G)_{12}$ / $(G)_{11 \text{ or } 10}$	7	6
4	SNPG2 / SNPG2	22	27
5	SNPG2 / $(G)_{10}$	2	1
6	SNPG1 / SNPG1	11	10
7	SNPG2 / SNPG6	8	5
8	SNPG4 / SNPG4	2	1
9	SNPG4 / ND*	4	3
10	SNPG4 / SNPG1	0	1
11	SNPG4 / $(G)_{10}$	0	1
12	SNPG5 / SNPG5	3	4
13	SNPG1 / SNPG2	1	3
14	$(G)_{13}$ / $(G)_{12}$	0	1
15	SNPG2 / SNPG4	3	5
16	SNPG1 / $(G)_{11}$	0	3
Not assigned		**1**	**2**
Fail PCR		**6**	**16**
Total		**142**	**145**

***ND**: second allele Not Determined.

good quality sequence data was obtained from repeat sequencing then this information was used to determine genotypes. If the sequencing data was consistently poor upon re-sequencing then the samples were subjected to the initial PCR amplification again (see method section 3.2.15.1). In the absence of good quality sequence data upon repeat of the initial PCR amplification a genotype was not determined for that sample.

Finally, to confirm further the genotypes identified a random batch of samples of each genotype was subject to repeat sequencing and these results were in agreement with those obtained originally. In addition, the integrity of the method used to manually interpret the sequencing chromatograms and assign a genotype which was adopted from Yu et al, was tested in a blind study. A random batch of direct sequencing data for each genotype was selected and using the guidelines as described previously (see section 8.4.1.2) the genotype of each sample was determined independently by a second individual. This line of analysis showed that the results of the genotypes determined independently by a second individual were in complete agreement with those determined originally. Hence, confirming the reliability of the method used to manually interpret discrete poly-G alleles from direct sequencing chromatograms and the accuracy of the genotypes determined, particularly in the case of complex

heterozygotes. Given the complexity of the -3432poly-G polymorphism it was reasoned that an alternative PCR to genotype the -3432poly-G polymorphism by direct sequencing using PCR and sequencing primers different from those used originally would be an appealing approach for further validation of the original data. Especially since sizing poly-G alleles by PAGE analysis of radioactive PCR products would not be able to rule out loss of heterozygosity in the primary screen by direct sequencing due to a mutation within the annealing site of primer 3cP12-AS leading to a 'null' allele. Consequently, attempts were made to optimise an alternative PCR reaction using primers 3cP12-S and G3cP13-AS for the initial PCR reaction and primer G3cP13-AS for sequencing (see Table 8.4, page 282) but, this approach was besieged with a high background upon direct sequencing and so not pursued further.

In summary, a total of 16 different -3432poly-G genotypes were identified from genotyping samples for the preliminary association study (see Table 8.6). An overview of the steps taken that lead to confirmation of each genotype is shown in Fig. 8.7.

8.4.4 Identification of haplotypes for -3495$(GT)_n$, -3457$(GA)_n$ and -3432poly-G polymorphism.

From the initial PCR-SSCP screen and genetic analysis to genotype the -3432poly-G polymorphism it became apparent that the adjacent dinucleotide repeats: -3457$(GA)_n$ and -3495$(GT)_n$ appear in strong linkage disequilibrium with the -3432poly-G polymorphism. This was particularly clear from analysis of subjects homozygous for the -3432poly-G polymorphism where sequencing of cloned PCR products for -3432poly-G polymorphism in combination with direct sequencing analysis revealed some very distinct haplotypes. For example, it was strikingly evident that poly-G allele SNPG1, one of the shortest poly-G alleles identified was always detected in combination with the shortest $(GT)_n$ dinucleotide repeat identified at position -3495; 15 repeats and the -3457$(GA)_n$ repeat haplotype detected in these samples always consisted of two repeats (see Fig. 4.14a, page 157). In contrast, poly-G allele SNPG2 was always detected in conjunction with a much longer $(GT)_n$ repeat at position -3495 (20 or 21) and the common -3457$(GA)_n$ haplotype observed in these samples was two repeats (see Fig. 4.14d, page 157). Similarly, if poly-G allele SNPG4 was present at position -3432 the native -3457 and -3495 haplotype detected was always $(GA)_1$ and $(GT)_{16}$; a -3457$(GA)_n$ repeat greater than one was never detected from the genetic analysis performed (see Fig. 4.14b, page 157).

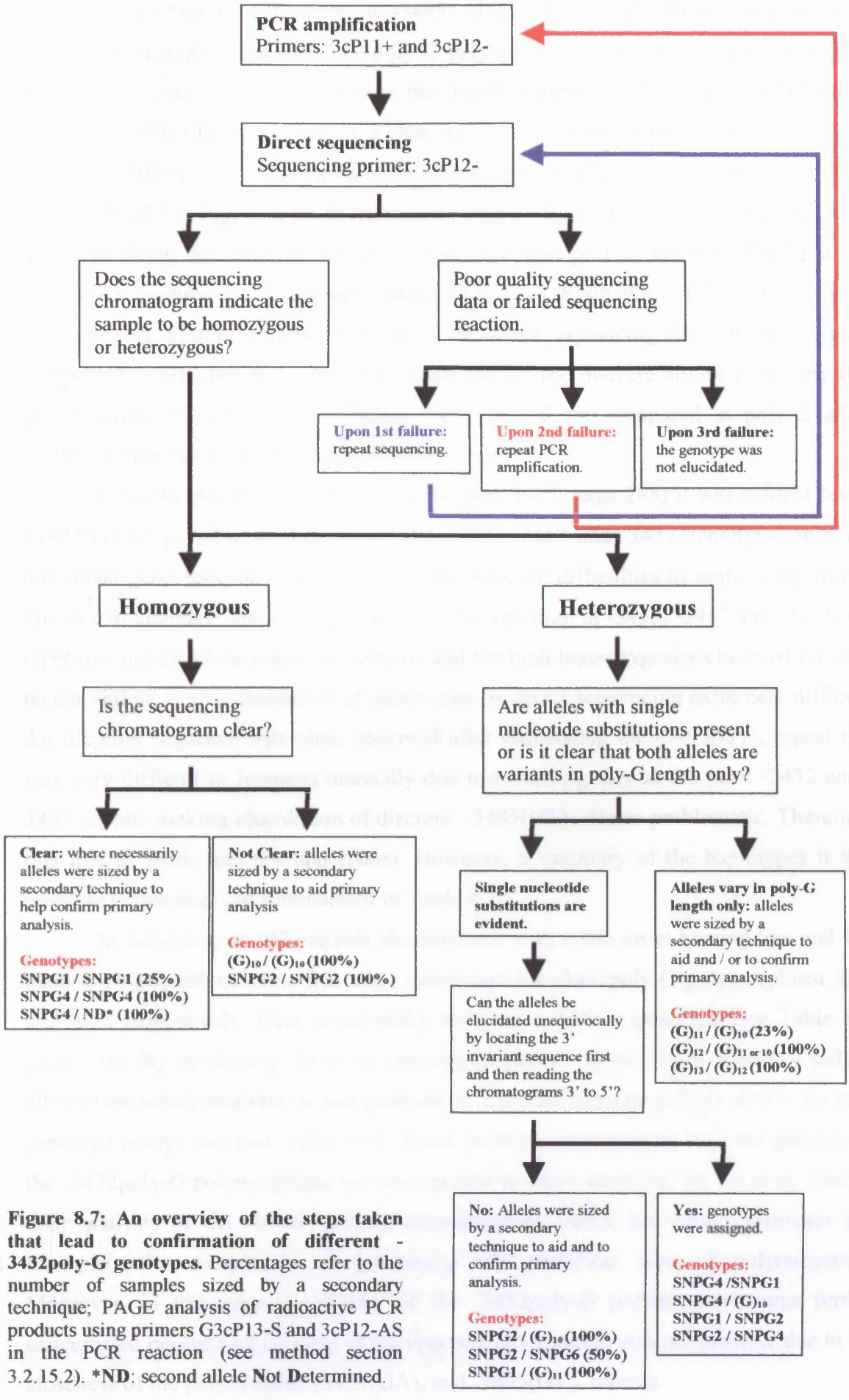


Figure 8.7: An overview of the steps taken that lead to confirmation of different - 3432poly-G genotypes. Percentages refer to the number of samples sized by a secondary technique; PAGE analysis of radioactive PCR products using primers G3cP13-S and 3cP12-AS in the PCR reaction (see method section 3.2.15.2). ***ND**: second allele **N**ot **D**etermined.

Elucidation of haploypes at -3495, -3457 and -3432 although mainly from subjects homozygous for the -3432poly-G polymorphism was by no means limited to just homozygotes. Figure 8.8a shows the direct sequencing chromatogram typical of SNPG2 / SNPG4 heterozygotes. It is clear from this chromatogram that the -3457$(GA)_n$ haplotype differs by one repeat between the two poly-G alleles present (shown as 'TC' versus 'TCTC' in Fig. 8.8a for the anti-sense strand). By locating the invariant sequence and elucidating the discrete alleles it is evident that poly-G allele SNPG2 has a -3457$(GA)_n$ haplotype of two repeats and poly-G allele SNPG4 a -3457$(GA)_n$ haplotype of one repeat. Similarly, figure 8.8b shows the direct sequencing chromatogram typical of SNPG2 / $(G)_{10}$ heterozygotes. By distinguishing the discrete alleles it is clear that poly-G allele $(G)_{10}$ has a -3457$(GA)_n$ haplotype of one compared to poly-G allele SNPG2 which has a -3457$(GA)_n$ haplotype of two.

From the preliminary genotyping (section 8.4.2, page 285) it was evident that it would not be possible to characterise the -3495, -3457 and -3432 haplotypes in every individual genotyped for -3432poly-G polymorphism; difficulties in sequencing limited this line of analysis. The close proximity of the variation at -3432, -3457 and -3495, the repetitive nature of the sequence variants and the high heterozygosity observed for each repeat variant made elucidation of haplotypes by direct sequencing extremely difficult. An illegible sequence was often observed after sequencing the -3495$(GT)_n$ repeat that was very difficult to interpret manually due to heterozygosity at the prior -3432 and -3457 repeats making elucidation of discrete -3495$(GT)_n$ alleles problematic. Therefore, this line of work was not undertaken. However, a summary of the haplotypes it was possible to decipher are summarised in Table 8.7.

In summary, of 142 patients characterised with a late onset hearing loss and 145 general population controls that were genotyped for -3432poly-G polymorphism 95% and 88% respectively, were successfully assigned a distinct genotype (see Table 8.6, page 291). By combining direct sequencing (primary analysis) with sizing of poly-G alleles (secondary analysis) it was possible to elucidate discrete poly-G alleles for each genotype except two (see Table 8.6). These problems encountered with the genotyping the -3432poly-G polymorphism are very similar to those described by Yu et al, 2002 in their analysis of the similar poly-G polymorphism within the KLK1 promoter and exemplify the complexity of genotyping this particular type of polymorphism. Moreover, in this case genotyping of the -3432poly-G polymorphism was further exacerbated because sequencing of the opposite DNA strand was not possible due to the closeness of the polymorphic -3457$(GA)_n$ and -3495$(GT)_n$ repeats.

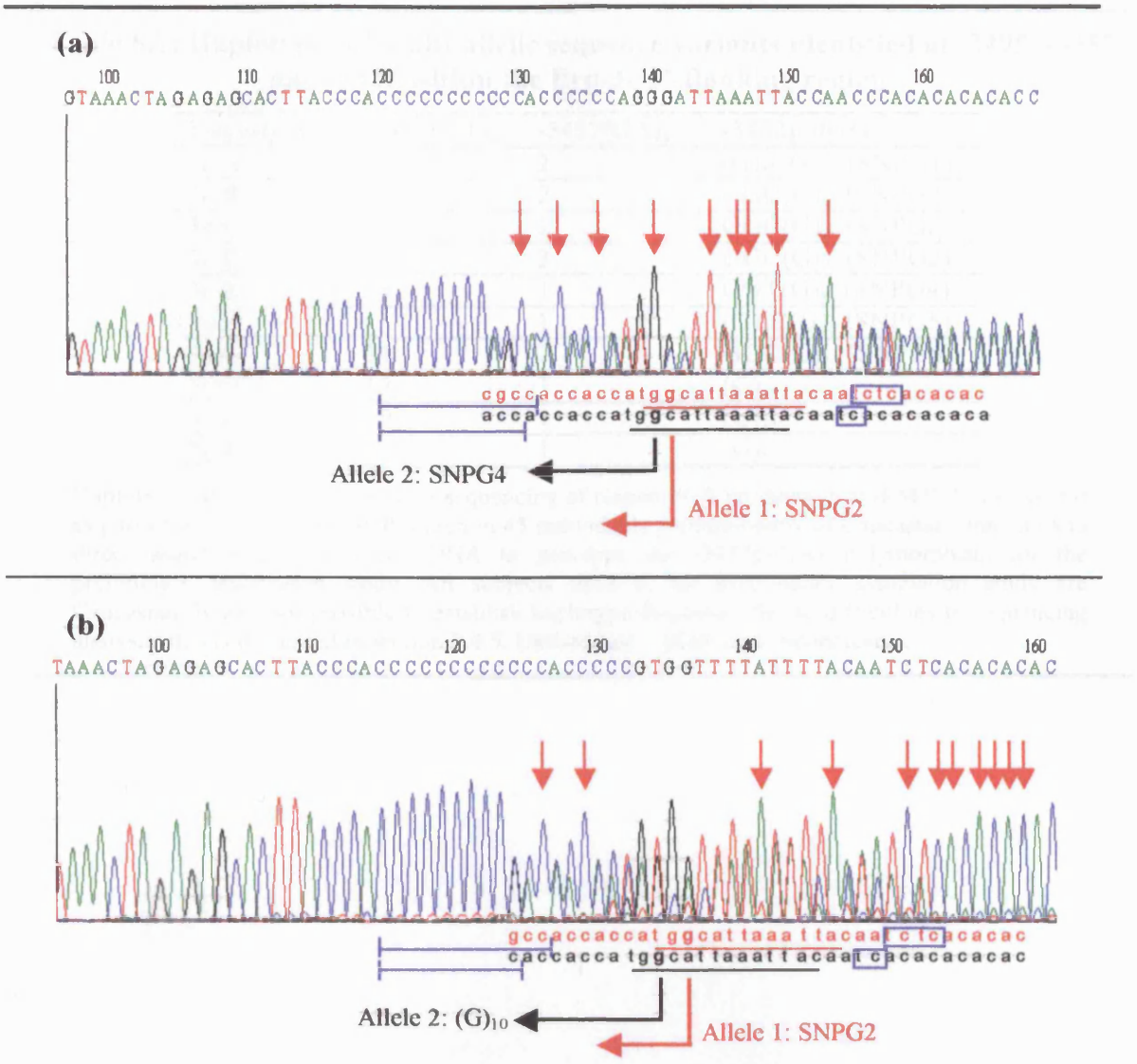


Figure 8.8 Direct sequencing chromatograms representative of individuals identified SNPG2 heterozygous. **(a)** and **(b)** The direct sequencing data obtained is for the anti-sense strand. Alleles were elucidated by reading 3' to 5' from the 3' G within the invariant region (underlined). Clear distinction between homozygous and heterozygous peaks facilitated identification of each allele; see written sequence added below each chromatogram (homozygous peaks are indicated by the small red arrows). Haplotypes for -3432poly-G polymorphism and -3457$(GA)_n$ repeat can be observed on each allele. The region of the sequencing chromatogram representing the -3432poly-G polymorphism is underlined in blue and boxed in blue for the -3457$(GA)_n$ repeat within the written sequence added below each chromatogram. **(a)** Direct sequencing chromatogram representative of subjects identified SNPG2 / $(G)_{10}$ heterozygous at -3432. The haplotype at -3432 and -3457 is SNPG2 : $(GA)_2$ respectively, for allele 1 and SNPG4 : $(GA)_1$ respectively, for allele 2. **(b)** Direct sequencing chromatogram representative of subjects identified SNPG2 / SNPG4 heterozygous at -3432. The haplotype at -3432 and -3457 is SNPG2 : $(GA)_2$ respectively, for allele 1 and $(G)_{10}$: $(GA)_1$ respectively, for allele 2.

Table 8.7: Haplotypes of multi-allelic sequence variants identified at -3495, -3457 and -3432 within the Brn-3c 5'-flanking region.

Haplotype	-3495$(GT)_n$	-3457$(GA)_n$	-3432poly-G
3c/A	15	2	$(G)_8CG$ (SNPG1)
3c/B	20	2	$GGC(G)_9$ (SNPG2)
3c/C	21	2	$GGC(G)_9$ (SNPG2)
3c/D	20	3	$GGC(G)_9$ (SNPG2)
3c/E	16	1	$GGT(G)_8$ (SNPG4)
3c/F	-	1	$(G)_4C(G)_6$ (SNPG5)
3c/G	24	2	$GGC(G)_{10}$ (SNPG6)
3c/H	19	1	$(G)_{10}$
3c/I	19	1	$(G)_{11}$
3c/J	-	1	$(G)_{12}$

Haplotypes identified are based on sequencing of cloned PCR products in pGEM®-T easy vector as part of an initial PCR-SSCP screen in 45 individuals predominantly of Caucasian ethnicity and direct sequencing of genomic DNA to genotype the -3432poly-G polymorphism for the preliminary association study. All subjects used in the preliminary association study are Caucasian. It was not possible to establish haplotype frequency due to difficulties in sequencing analysis; this is discussed in section 8.4.5. Dashed line – allele type inconclusive.

8.5 Association analysis to asses whether the -3432poly-G polymorphism is a risk factor for late onset hearing loss.

To assess whether the -3432poly-G polymorphism is a risk factor for late onset hearing loss the -3432poly-G polymorphism was genotyped in our late onset hearing loss and general population cohorts for a preliminary association study. It was accepted at the outset of this research project that the number of patients characterised with late onset hearing loss that could be recruited into this study within the time frame of the PhD project would limit the power of the association analysis. Therefore, the main aim of genotyping the -3432poly-G polymorphism for this PhD project was to establish allele frequency and to perform preliminary association analysis as a first step prior to undertaking a large-scale population based case-control association study, which is beyond the capacity of the PhD project.

Due to the complexity of the -3432poly-G polymorphism: multiple variations in length of the poly-G repeat coupled with various single nucleotide substitutions within the poly-G repeat, a dual approach was devised for genotyping. Direct sequencing was used as the primary analysis and sequencing chromatograms were interpreted manually using an adaptation of Yu et al, 2002. Sizing of poly-G alleles by PAGE analysis of radioactive PCR products was used in secondary analysis to aid elucidation of discrete poly-G alleles in complex heterozygotes and to confirm data obtained by direct sequencing.

Using this approach 95% of patients and 88% of general population samples genotyped for -3432poly-G polymorphism were assigned a specific genotype (see Table 8.8). A total of 16 distinct genotypes were identified and poly-G alleles were elucidated for each genotype except two. For subjects with genotype $(G)_{12}$ / $(G)_{11\ or\ 10}$ it was not possible to discriminate between $(G)_{11}$ and $(G)_{10}$ as the second allele type. Whereas, for subjects with genotype SNPG4 / ND it was not possible to clarify the second allele type. In both these cases data obtained by PAGE analysis of radioactive PCR products did not prove useful when coupled with data obtained by direct sequencing (a full discussion of these problems encountered in genotyping the -3432poly-G polymorphism is given in Appendix H: *elucidation of problematic genotypes*). Therefore, allele frequencies were determined accommodating these two limitations. For subjects identified as $(G)_{12}$ / $(G)_{11\ or\ 10}$ both possibilities were counted as the second allele type and for subjects identified SNPG4 / ND this category was excluded from subsequent frequency analysis. It was

Table 8.8: Genotype frequency of -3432poly-G polymorphism amongst Caucasian individuals in late onset hearing loss and general population cohorts.

	Late onset hearing loss		General population	
Genotype	**Observed**	**Frequency**	**Observed**	**Frequency**
$(G)_{11}$ / $(G)_{10}$	69	0.527	52	0.419
$(G)_{10}$ / $(G)_{10}$	3	0.023	4	0.032
$(G)_{12}$ / $(G)_{11\ or\ 10}$	7	0.053	6	0.048
SNPG2 / SNPG2	22	0.168	27	0.218
SNPG2 / $(G)_{10}$	2	0.015	1	0.008
SNPG1 / SNPG1	11	0.084	10	0.081
SNPG2 / SNPG6	8	0.061	5	0.040
SNPG4 / SNPG4	2	0.015	1	0.008
SNPG4 / ND*	4	-	3	-
SNPG4 / SNPG1	0	0.000	1	0.008
SNPG4 / $(G)_{10}$	0	0.000	1	0.008
SNPG5 / SNPG5	3	0.023	4	0.032
SNPG1 / SNPG2	1	0.008	3	0.024
$(G)_{13}$ / $(G)_{12}$	0	0.000	1	0.008
SNPG2 / SNPG4	3	0.023	5	0.040
SNPG1 / $(G)_{11}$	0	0.000	3	0.024
Not assigned	**1**		**2**	
Fail PCR	**6**		**16**	
Total	**142**		**145**	

*ND: second allele Not Determined.

acknowledged that this approach may lead to an over-representation of alleles $(G)_{10}$ and $(G)_{11}$, and an under-representation of allele SNPG4 in the late onset hearing loss sample group examined. However, since the same criteria was applied to the general population sample group and similar numbers of samples were identified as $(G)_{12}$ / $(G)_{11\ or\ 10}$ or SNPG4 / ND between both sample groups the possibility of these limitations leading to a false-positive association is unlikely.

From Table 8.8 it is evident that there is a skew towards certain genotypes. The most common genotype in both the late onset hearing loss and general population sample groups is $(G)_{11}$ / $(G)_{10}$ heterozygous closely followed by SNPG2 homozygous. $(G)_{11}$ / $(G)_{10}$ heterozygotes account for around 40-50% of all genotypes identified within each sample group. Given the high frequency of the $(G)_{11}$ / $(G)_{10}$ genotype across both sample groups the $(G)_{11}$ and $(G)_{10}$ homozygotes appear under-represented (Table 8.8). Indeed, the $(G)_{11}$ homozygous genotype was not identified during the course of this analysis raising the possibility that some samples identified $(G)_{11}$ / $(G)_{10}$ heterozygous may have been wrongly identified. However, samples $(G)_{11}$ / $(G)_{10}$ heterozygous exhibit very clear and distinctive results upon direct sequencing as illustrated by the chromatogram shown in Fig. 8.5 (page 284) and it is unlikely that any have been miss-

assigned on sequencing chromatogram alone. In addition, in a proportion of subjects (23%) assigned genotype $(G)_{11}$ / $(G)_{10}$ allele size was determined by PAGE analysis of radioactive PCR products and this was found to be in complete agreement with the original data obtained by direct sequencing (see Fig. 8.7 page 292). Hence, confirming the original genotype identified.

Due to the high proportion of subjects identified SNPG2 homozygous (17% and 22% in late onset hearing loss and general population sample groups, respectively) measures were undertaken to help exclude the possibility of loss of heterozygosity in the primary screen by direct sequencing. Allele size was determined by PAGE analysis of radioactive PCR products in each sample identified from direct sequencing to be SNPG2 homozygous (see Fig. 8.7 page 292) as this would help exclude the possibility of non-amplification of one allele in the primary screen by direct sequencing due to stochastic hypoamplification leading to allele drop-out. This line of secondary analysis confirmed the original SNPG2 homozygous genotype assignment in every sample analysed except three; for three samples the presence of an extra band upon PAGE analysis of radioactive PCR products was observed that was too intense to be regarded an artefact from the PCR reaction. It was not possible to clarify the genotype of these ambiguous samples; they were left unassigned and excluded from subsequent frequency analysis (see Table 8.8). Since similar numbers of ambiguous samples are observed between both late onset hearing loss and general population sample groups one and two samples respectively, this is unlikely to have an affect on subsequent allele frequency analysis.

It should also be remembered that many additional measures were undertaken to confirm all distinct poly-G genotypes identified. In addition to sizing of poly-G alleles by PAGE analysis of radioactive PCR products repeat sequencing and a blind study was undertaken and the results of this analysis on random batches of samples for each genotype including $(G)_{11}$ / $(G)_{10}$ and SNPG2 homozygous were in complete agreement with the original genotypes identified. Thus, further confirming the accuracy of the methods undertaken to elucidate discrete poly-G alleles and the reliability of the original data.

A total of nine discrete alleles were identified for the -3432poly-G polymorphism using our late onset hearing loss and general population cohorts of Caucasian ethnicity and allele frequencies are shown for both cohorts in Table 8.9. Alleles $(G)_{10}$, $(G)_{11}$, SNPG2 and SNPG1 are the most common alleles observed in both late onset hearing loss and

Table 8.9: Allele frequency of -3432poly-G polymorphism amongst Caucasian individuals in late onset hearing loss and general population cohorts

		Late onset hearing loss		Control	
Allele	**Affinity of allele for SP1**	Observed	Frequency	Observed	Frequency
SNPG1	High	23	0.086	27	0.106
SNPG2	Low	58	0.216	68	0.268
SNPG4	Low	7	0.026	9	0.035
SNPG5	Intermediate	6	0.022	8	0.031
SNPG6	-	8	0.030	5	0.020
$(G)_{10}$	High	84	0.312	68	0.268
$(G)_{11}$	Intermediate	76	0.283	61	0.240
$(G)_{12}$	Intermediate	7	0.026	7	0.028
$(G)_{13}$	-	0	0.000	1	0.004
Total		269		254	

Alleles were grouped according to affinity for SP1: high affinity (red), low affinity (blue) and intermediate affinity (green) based on the ability of each poly-G allele to compete with the consensus SP1 binding site for recombinant SP1 protein in EMSA analysis (see Fig. 5.7 section 5.3.2). Dashed line: allele was not included in EMSA analysis and affinity of allele for SP1 is unknown.

general population sample groups accounting for a similar proportion of alleles in both sample groups: 90% and 80% of total alleles, respectively. The remaining five alleles: SNPG4, SNPG5, SNPG6, $(G)_{12}$ and $(G)_{13}$ are much rarer in both sample groups. Although, the same poly-G alleles are either common or rare in both sample groups it is clear that slight differences in allele frequency exist between these two sample groups (see Table 8.9).

In-vitro functional characterisation of the -3432poly-G polymorphism shows that the -3432poly-G polymorphism modifies a high affinity SP family binding site and there is convincing evidence that this alters the activity of the Brn-3c promoter when SP1 levels are limiting (see Chapter 5, sections 5.3 – 5.5). Interestingly, allele SNPG2 which contains a low affinity SP1 binding site is more prevalent in the general population sample group whereas, allele $(G)_{10}$ which contains a high affinity SP1 binding site is more prevalent in the late onset hearing loss sample group. However, the other high affinity SP1 binding allele, SNPG1, does not follow this trend being slightly more prevalent in the general population sample group compared to the late onset hearing loss sample group. Therefore, from these observations there is no clear pattern between affinity of poly-G allele for SP1 and allele prevalence between patients characterised with late onset hearing loss compared to a general population sample.

Although, no clear trend is apparent between affinity of poly-G allele for SP1 and allele prevalence between patients characterised with late onset hearing loss compared to a general population sample a Chi (χ^2) square test was performed to assess

whether the observed differences in poly-G allele frequency between these two sample groups is significant (see method section 3.2.17). The results of this analysis did not reveal a significant difference in allele frequency between the two sample groups; the slight differences in poly-G allele frequency observed between late onset hearing loss and general population cohorts are too small to reach statistical significance in this sample size. Hence, suggesting that there is no association between -3432poly-G allele, high affinity SP1 binding allele or otherwise and susceptibility to late onset hearing loss, *at least in the context of this preliminary association study*. This result is discussed further in section 8.8: General discussion.

8.6 Genotyping the -1391A>C single nucleotide substitution.

The -1391A>C sequence variant in the Brn-3c promoter is a single nucleotide substitution. At the time of commencing this project the -1391A >C single nucleotide substitution had been reported in the public NCBI SNP database (NCBI SNP Cluster ID: rs1368402), but had not been validated and no report of allele frequency was available. Mutation scanning of the Brn-3c gene by PCR-SSCP analysis to identify novel sequence variants at the Brn-3c locus confirmed the -1391 single nucleotide substitution and suggested that this sequence variant is common in the general population with a rare allele frequency of 0.28 (based on SSCP analysis on 45 individuals; see Table 4.2, section 4.3.1, page 146). However, PCR-SSCP analysis is not a reliable method for genotyping as has been discussed previously (section 4.3.1, page 144) and therefore, the rare allele frequency of 0.28 was taken as an estimate only. Hence, an alternative methodology for genotyping the -1391A>C single nucleotide substitution was sought.

One hundred and forty-two patients with late onset hearing loss and 145 general population samples were available for genotyping in a preliminary association study during the course of this project. Given the moderate size of the sample cohorts it was decided to assess whether the -1391A>C single nucleotide substitution in the Brn-3c promoter alters the recognition sequence for a restriction endonuclease for genotyping by PCR-restriction fragment length polymorphism (RFLP) analysis (for review see Gut, 2001). PCR-RFLP analysis is a traditional and one of the most commonly used methods for genotyping single nucleotide substitutions in standard laboratories. Restriction endonucleases, generally known as restriction enzymes recognise short, specific, often palindromic double stranded DNA sequences and cleave the double stranded DNA either within or adjacent to the recognition site. Typically, a PCR fragment spanning the single nucleotide substitution is amplified and subjected to restriction enzyme digest using a restriction enzyme whose recognition site is modified by the single nucleotide substitution. This results in restriction site cleavage only for the allele that does not change the recognition site. Consequently, samples can be genotyped by observation of the restriction fragments produced upon agarose gel electrophoresis. A small sequence of DNA spanning the -1391 sequence variant was scanned against the known restriction enzyme sites using Webcutter software version 2.0 (www.firstmarket.com/cutter/cut2.html). The results of this analysis revealed that the -

1391 sequence variant modifies the restriction site recognised by *TaqI*. *TaqI* recognises the sequence 'TCGA' and cleaves double stranded DNA within this site; the variant allele, -1391C modifies the *TaqI* site to 'TCGC' preventing recognition by *TaqI* (Fig. 8.9a). Therefore, it was decided to genotype the -1391 single nucleotide substitution by PCR-*TaqI* digest for a simple, cost-effective and speedy approach.

8.6.1 Genotyping the -1391 single nucleotide substitution by *TaqI* digest.

In order to accurately genotype the -1391A>C single nucleotide substitution by *TaqI* digest the main priority was to amplify by PCR a suitably sized fragment to ensure good size separation of bands upon agarose gel electrophoresis for reliable allocation of genotypes. In addition, to make certain assignment of the homozygous variant genotype was not due to failure of *TaqI* to cleave the amplicon in any particular sample (*TaqI* does not recognise the DNA sequence produced when the variant allele, -1391C is present) it was essential that the fragment amplified included an additional *TaqI* site as an internal control. With these features in mind a PCR reaction was optimised using primers 3cP1-S: 5' CTCTCAGCGGAGGCAGTGG 3' and 3cP3-AS: 5' CCGTCTAAGGAAGCTTGTGGAG 3' originally designed for the PCR-SSCP screen of the Brn-3c promoter (details of the PCR optimisation and subsequent *TaqI* digest are described in method section 3.2.16). Primers 3cP1-S and 3cP3-AS generate a 596bp product upon PCR within which *TaqI* cleaves a constant *TaqI* site at position 204bp and the wild-type but not the variant allele, -1391C at position 81bp (see Fig. 8.9b and c). Accordingly, upon agarose gel electrophoresis subjects homozygous for the wild-type allele, -1391A should generate three bands of 81, 123 and 392bp, subjects homozygous for the variant allele, -1391C two bands of 204 and 392bp, and subjects heterozygous for the -1391 single nucleotide substitution all four bands: 81, 123, 204 and 392bp (see Fig. 8.9b and c).

Having devised and optimised a suitable protocol for genotyping the -1391 single nucleotide substitution by *TaqI* digest the -1391 single nucleotide substitution was genotyped in 142 patients characterised as having a late onset sensorineural hearing loss and 145 general population samples for a preliminary association study. In the absence of a banding pattern for any sample or very faint bands, the PCR-*TaqI* digest was repeated and the genotype assigned accordingly. If absence of bands persisted and no PCR product was visualised upon repeat PCR this was regarded a PCR failure and no genotype could be assigned. An example of the banding pattern produced upon

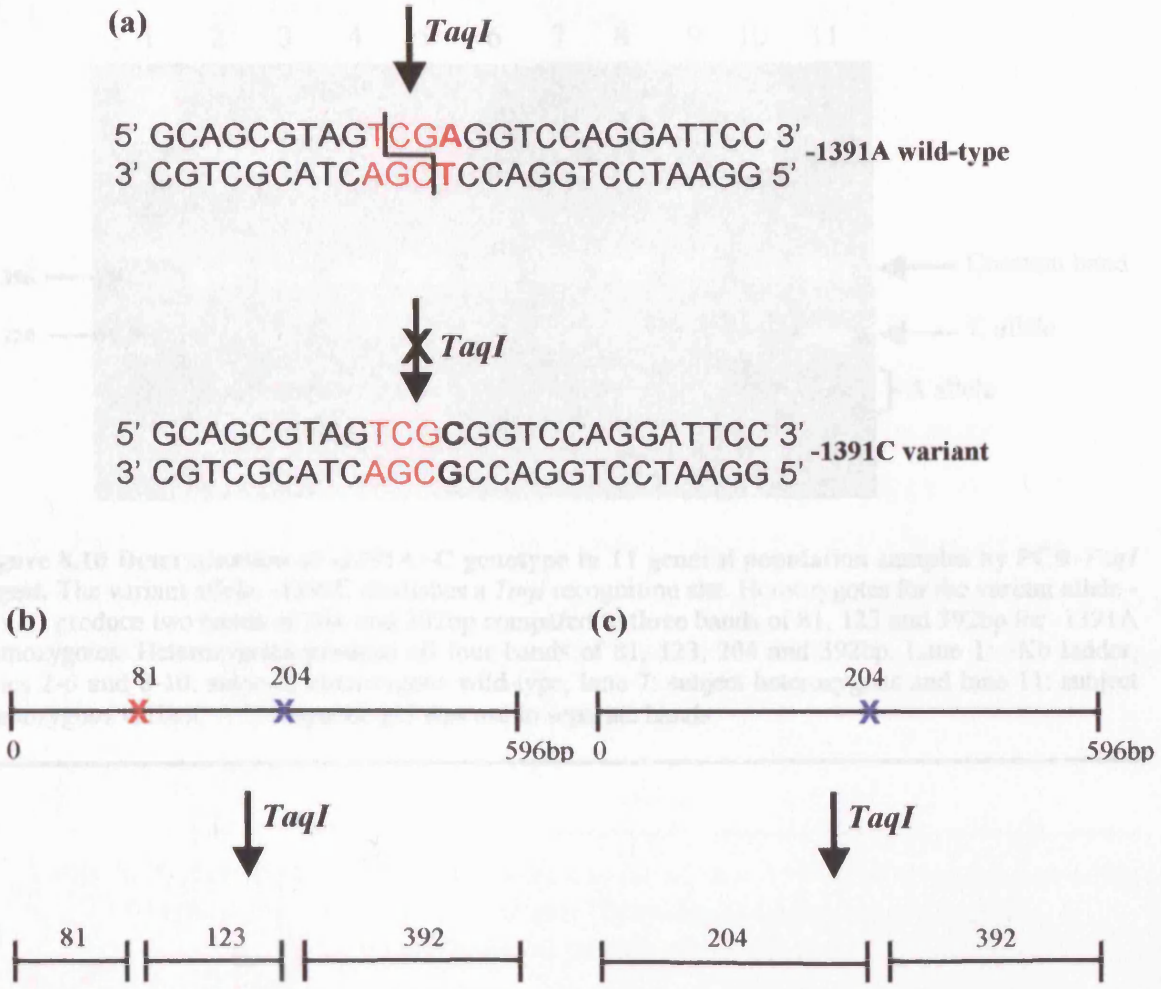


Figure 8.9: *TaqI* recognises the wild-type allele -1391A but not the variant allele, -1391C. (a) Section of the Brn-3c promoter spanning the -1391A>C sequent variant to show the wild-type allele, A and variant allele, C. The wild-type allele, A creates a *TaqI* recognition site (highlighted in red with the wild-type allele in red bold) and *TaqI* cleaves double stranded DNA within this site (shown by staggered black lines). The variant allele, C (shown in black bold) abolishes the *TaqI* recognition site preventing cleavage by *TaqI*. **(b)** and **(c)** Schematic of 596bp PCR product produced using primers 3cP1-S and 3cP3-AS. Red cross denotes *TaqI* recognition site formed by presence of the wild-type allele -1391A, which is located at position 81bp within the 596bp fragment. Blue cross denotes a constant *TaqI* recognition site within the 596bp fragment used as an internal control in *TaqI* restriction digest for activity of *TaqI*. **(b)** Shows the *TaqI* restriction fragments produced for subjects homozygous for the wild-type allele, -1391A. **(c)** Shows the *TaqI* restriction fragments produced for subjects homozygous for the variant allele, -1391C (for clarity only one allele is shown in b and c). Subjects heterozygous at -1391 will produce all four bands of 81, 123, 204 and 392bp.

agarose gel electrophoresis following *TaqI* digest is shown in Fig. 8.10 and the results obtained and subsequent association analysis performed is discussed in the next section.

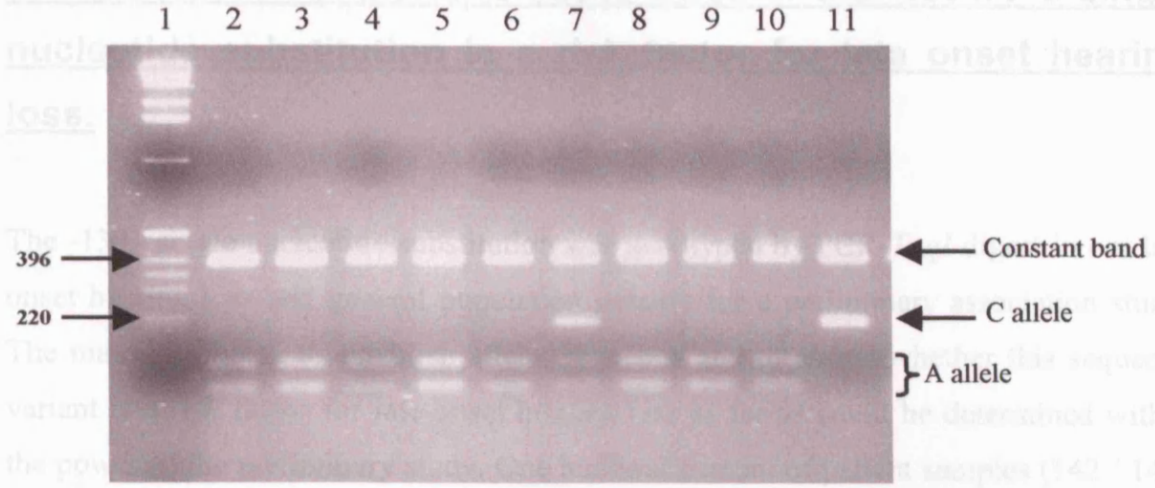


Figure 8.10 Determination of -1391A>C genotype in 11 general population samples by PCR-*TaqI* digest. The variant allele, -1391C abolishes a *TaqI* recognition site. Homozygotes for the variant allele -1391C produce two bands of 204 and 392bp compared to three bands of 81, 123 and 392bp for -1391A homozygotes. Heterozygotes produce all four bands of 81, 123, 204 and 392bp. Lane 1: 1Kb ladder, lanes 2-6 and 8-10: subjects homozygous wild-type, lane 7: subject heterozygous and lane 11: subject homozygous variant. A 2% agarose gel was use to separate bands.

8.7 Association analysis to asses whether the -1391A>C single nucleotide substitution is a risk factor for late onset hearing loss.

The -1391 single nucleotide substitution was genotyped by PCR-*TaqI* digest in our late onset hearing loss and general population cohorts for a preliminary association study. The main aim being to establish allele frequency and to assess whether this sequence variant is a risk factor for late onset hearing loss as far as could be determined within the power of the preliminary study. One hundred percent of patient samples (142 / 142) and 94% (136 / 145) of general population samples were successfully genotyped for the -1391 sequence variant (see Table 8.10). In a few general population samples (6%) it was not possible to identify genotype due to repeated PCR failure and therefore, these samples were excluded from subsequent frequency analysis.

Table 8.10: Genotype and allelic frequencies of the -1391A>C single nucleotide substitution amongst Caucasian individuals in late onset hearing loss and general population cohorts.

	Genotype	Observed	Genotype frequency	Allele	Allele frequency
Late onset hearing loss	**-1391**	**142**		A	0.78
	AA	86	0.61	C	0.22
	AC	49	0.34		
	CC	7	0.05		
General population	**-1391**	**136**		A	0.71
	AA	73	0.54	C	0.29
	AC	47	0.34		
	CC	16	0.12		

To assess whether the genotype distribution for the -1391A>C single nucleotide substitution is in Hardy-Weinberg equilibrium in the late onset hearing loss and general population sample groups, predicted and observed genotypes were compared by χ^2 analysis (see method section 3.2.17). In the late onset hearing loss sample group predicted values for all three genotypes assuming Hardy-Weinberg equilibrium are AA: AC: CC = 86.39: 48:73: 6.87 and actual observed values are AA: AC: CC = 86: 49: 7.

Whereas, for the general population sample group predicted values assuming Hardy-Weinberg equilibrium are AA: AC: CC = 68.47: 56.06: 11:47 and actual observed values are AA: AC: CC = 73: 47: 16. These observed and predicted values do not differ significantly by χ^2 analysis in either the late onset hearing loss or general population sample groups and hence, show that both the late onset hearing loss and general population sample groups are in Hardy-Weinberg equilibrium with regard to the distribution of the -1391A>C single nucleotide substitution.

From *in-vitro* functional characterisation of the -1391 single nucleotide substitution it is evident that the variant allele, C has a reduced affinity for an OC-2 derived nuclear protein and this is consistent with a subtle but, significant decrease in basal activity of the Brn-3c promoter (section 7.2 and 7.5). Thus, it is tempting to speculate that possession of the variant C allele could lead to reduced expression of Brn-3c and increase the genetic risk for late onset hearing loss. However, from Table 8.10 it is clear that the homozygous genotype for the variant allele, -1391C, is more prevalent in the general population sample group compared to the late onset hearing loss sample group; comparing a homozygous (CC) genotype frequency of 0.12 versus 0.05, respectively. Hence, suggesting that possession of the homozygous variant genotype (CC) is associated with reduced rather than increased risk of late onset hearing loss, which is clearly not consistent with the hypothesis that possession of the variant C allele could lead to reduced expression of Brn-3c and increase the genetic risk for late onset hearing loss. One explanation for this anomaly could be that the -1391A>C single nucleotide substitution is not causative for late onset hearing loss but in linkage disequilibrium with a causative variant for late onset hearing loss. However, when examined by χ^2 analysis the difference in genotype frequencies between late onset hearing loss and general population sample groups is too small to be statistically significant. Similarly, the difference in allele frequency between the late onset hearing loss and general population sample groups (in the general population sample group the rare -1391C allele frequency is 0.29 and this is slightly higher than a frequency of 0.22 observed for the late onset hearing loss sample group; see Table 8.10) is not statistically significant when examined by χ^2 analysis. Thus, collectively this data suggests that there is no association between the -1391A>C single nucleotide substitution and susceptibility to late onset hearing loss, *at least in the context of this preliminary association study*. This result is discussed further in the next section, 8.8: General discussion.

8.8 General discussion.

The preliminary association analysis performed for this thesis did not find an association between either the -3432poly-G polymorphism or the -1391A>C single nucleotide substitution and susceptibility to late onset hearing loss. The subtle differences in allele frequency observed for both the -3432poly-G polymorphism and the -1391A>C single nucleotide substitution were not significantly different in our late onset hearing loss and general population cohorts when examined by χ^2 analysis. However, this result is not unexpected and it is important to consider the nature of late onset hearing loss, which is a complex disease and the design of the association study when interpreting these results.

The design of an association study is an important determinant in finding a positive association and the likelihood of finding a positive association is influenced by many parameters. One of the most critical factors in good study design is the power of the study to detect an association, which is strongly dependent upon the size of the sample used; this is especially important when the relative risk carried by the causal allele is small and the frequency of the allele in the general population is low (for review see Donahue and Allen, 2005). The sample size needed to detect a genetic effect is related to the frequency and relative risk carried by the allele under study. In principle, if a gene under study has a large effect on disease phenotype the causal allele will carry a high relative risk and has the potential to be detected even when the size of the sample is small and / or when the frequency of the allele is low (for review see Risch, 2000; Donahue and Allen, 2005). However, late onset hearing loss is a complex disease many genes are thought to contribute to susceptibility and the relative risk contributed by any one gene is likely to be small. Consequently, subtle effects and genetic heterogeneity, coupled with interplay from the environment compounds the likelihood of finding a positive association in a *small-scale* pilot study such as the preliminary association study presented for this thesis.

It is also important to consider that the control sample group used to conduct preliminary association analysis for this thesis consisted of random members of the general population that on average were 40 years younger than the patient sample group. It was not possible to categorically exclude that these general population subjects have a late onset hearing loss or may go on to be affected by late onset hearing loss and this anomaly increases the likelihood of a false-negative association by reducing the

power of the association analysis (Moskvina et al, 2005). Furthermore, with respect to the -3432poly-G polymorphism it is important to recognise that this is not a simple dialleleic polymorphism; the large number of alleles identified (nine in total) further reduces the power of the association study, especially when only small numbers of samples are examined. Therefore, failure to find a positive association between either the -3432poly-G polymorphism or the -1391A>C single nucleotide substitution and susceptibility to late onset hearing loss within the context of this preliminary association study does not mean that these sequence variants have no affect on late onset hearing loss disease risk. *In-vitro* analysis suggests that the -3432poly-G polymorphism has an affect on regulation of the Brn-3c gene when SP1 levels are limiting, $p<0.05$ (paired t-test; see Fig. 5.13b, page 203) raising the possibility that this sequence variant could be causative for late onset hearing loss. Therefore, when all these factors are taken into consideration to categorically exclude that the -3432poly-G polymorphism is not associated with susceptibility to late onset hearing loss it is clear that a much larger cohort of samples than was possible to recruit during the course of this PhD project is required. Similarly, there is some evidence that the -1391A>C single nucleotide substitution has an affect on Brn-3c regulation; evidence suggests that this single nucleotide substitution has an effect, albeit subtle, on basal activity of the Brn-3c promoter, $p<0.05$ (t-test; Fig. 7.8b, page 257). Accordingly, it would be worthwhile extending this functional analysis in an attempt to gain more convincing evidence of the affect of the -1391 single nucleotide substitution on Brn-3c regulation. Stronger evidence for a functional effect of the -1391 single nucleotide substitution on Brn-3c regulation coupled with genotyping this sequence variant in a much larger cohort would allow one to conclude with more certainty the affect of this sequence variant on susceptibility to late onset hearing loss.

Certainly, use of a control sample group that has no medical history of hearing loss and that is more closely age-matched than the 'control' general population sample group used for the preliminary study would improve the power of the association analysis. With this in mind recruitment of age-matched cohorts characterised with late onset hearing loss and good hearing is now underway in our laboratory for use in future study. Samples are being recruited from two *Scottish Mental Health Survey Cohorts* as part of an on-going collaboration with Professor Lawrence Whalley (University of Aberdeen, Scotland, U.K.) and Dr Ian Deary (University of Edinburgh, Scotland, U.K.). These cohorts are a valuable resource for future case-control association analysis that is beyond the capacity of this PhD project. Recruitment of subjects into our late onset

hearing loss cohort from the adult hearing aid clinic at the Royal Free Nose, Ear and Throat Hospital, London, U.K. is also continuing with the aim to collect at least 1000 well-characterised samples. Both these London and Scottish late onset hearing loss cohorts pave the way for promising future analysis. In addition, because these cohorts are extremely well characterised and diagnostic audiogram is available these cohorts provide a resource by which to analyse late onset hearing loss as a quantitative trait (see Fransen et al, 2004; see also section 1.6.3). This approach avoids stringent 'case' versus 'control' classification of subjects. Instead, a continuous parameter for late onset hearing loss is defined and the *extent* by which a subject is effected by late onset hearing loss is used in subsequent statistical analysis in an attempt to find an association between sequence variants in a candidate gene and severity of late onset hearing loss. This is important as by treating late onset hearing loss as a quantitative trait as opposed to a binary trait this approach holds much promise as a power statistical method to detect susceptibility genes of moderate effect (Fransen et al, 2004).

Genotyping the -3432poly-G polymorphism and the -1391A>C single nucleotide substitution in our late onset hearing loss and general population cohorts established allele frequency (see Tables 8.9 and 8.10, respectively). The rare allele frequency for the diallelic -139A>C single nucleotide substitution in our general population cohort is 0.29 (Table 8.10). Based on this result it can be estimated that a sample size of approximately 600 would be needed to detect a small effect (relative risk: 1.6) with at least 80% power using a stringent level of significance (see Donahue and Allen, 2005). To detect a smaller effect (relative risk: 1.4) whilst maintaining the same level of power the sample size needed would increase to 1250. For the multi-allelic -3432poly-G polymorphism where the high allele number of variation identified (9 alleles) further limits the power of the association analysis it is likely that even larger sample sizes will be required.

Finally, it should be remembered that the data suggest the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism and native 5' haplotype is functional (see Chapter 6, section 6.3.2). Specifically, the data suggest functional interplay between the -566$(GT)_n$ repeat and the -3432poly-G polymorphism determines the response of the Brn-3c gene to SP1 (see Fig. 6.2, page 231). Therefore genotyping of the -3432poly-G polymorphism in combination with the -566$(GT)_n$ repeat would be necessary to determine whether specific haplotypes for these sequence variants in the Brn-3c 5'-flanking region are associated with increased risk of late onset hearing loss. If specific haplotypes for the poly-G polymorphism at -3432 and the (GT) dinucleotide

repeat at -566 in the Brn-3c promoter were elculcidated it may be that genotyping of only the -566$(GT)_n$ repeat would be necessary to determine if specific 5' haplotypes encompassing the -3432poly-G polymorphism and the -566$(GT)_n$ repeat are a risk factor for late onset hearing loss. This would be a practical approach as genotyping the -566$(GT)_n$ repeat in large sample cohorts would be technically more simple and less time consuming than genotyping the -3432poly-G polymorphism in such large cohorts.

8.9 Conclusion.

In conclusion, the data presented in this thesis is one of the first studies to identify and undertake a functional screen of common genetic variation in the Brn-3c gene prior to performing association analysis in an attempt to identify the genetic determinants that underlie late onset hearing loss. *In-vitro* analysis suggests that at least three common sequence variants in the Brn-3c promoter could have a functional effect: a complex poly-G repeat at -3432 that also exhibits single nucleotide substitutions within the poly-G repeat (poly-G polymorphism), a dinucleotide $(GT)_n$ repeat at -566 and a single nucleotide A to C substitution at -1391. Both the -3432poly-G polymorphism and the -1391A>C single nucleotide substitution were genotyped in a cohort of subjects characterised as having a late onset hearing loss alongside a similar number of general population samples to establish allele frequency and for a preliminary association study.

Subtle differences in allele frequency for both the -3432poly-G polymorphism and the -1391A>C single nucleotide substitution were found between the late onset hearing loss and general population sample groups examined, but were not significantly different when examined by χ^2 analysis (see section 8.5 and 8.7, respectively). However, this result is not unexpected when one considers the multi-factorial nature of late onset hearing, which is a complex disease coupled with some shortcomings in design of this preliminary association study (see section 8.8). Consequently, lack of a positive association between either the -3432poly-G polymorphism or the -1391A>C single nucleotide substitution and susceptibility to late onset hearing loss within the context of this preliminary association study can not be taken as conclusive evidence that these sequence variants are not a risk factor for susceptibility to late onset hearing loss. This is especially so in light of the evidence, which suggests that both these sequence variants could have a functional effect on Brn-3c regulation. Therefore, to conclude with certainty whether these sequence variants affect the genetic risk for late onset hearing loss further analysis that is beyond the scope of this PhD project is required. It is possible that by genotyping a much larger cohort of samples than was possible to recruit during the course of this PhD project or by performing quantitative trait analysis in a large late onset hearing loss cohort a positive association for either or both of these sequence variants will be established. Identification of a positive association between the -3432poly-G polymorphism and susceptibility to late onset hearing loss coupled with the evidence presented in this thesis which suggests that this

sequence variant has a functional effect on Brn-3c regulation together with additional functional analysis to exclude an affect of the -3495$(GT)_n$ repeat on Brn-3c promoter activity in response to SP1 would be more convincing evidence that this sequence variant modifies the genetic risk for late onset hearing loss. Replication of such an association in an independent cohort and / or a meta-analysis would validate this finding. This line of investigation would allow one to assess with more confidence the affect of the -3432poly-G polymorphism on susceptibility to late onset hearing loss. It would also be important to take into account the affect of the -566$(GT)_n$ repeat in such analysis as functional analysis suggests an association between the -566$(GT)_n$ repeat and the -3432poly-G polymorphism is important to Brn-3c regulation (see discussion in previous section, 8.8). Similarly, identification of a positive association between the -1391A>C single nucleotide substitution and susceptibility to late onset hearing loss, replication of this finding in an independent cohort coupled with more extensive functional analysis than is presented in this thesis would provide more convincing evidence that this sequence variant has an effect on susceptibility to late onset hearing loss and is not just in linkage disequilibrium with a causative variant for late onset hearing loss. These findings are summarised in the next Chapter: 9.0 General Discussion and Conclusion.

9.0 General Discussion and Conclusion.

The present study is one of the first to adopt a *functional screening approach* of *common* genetic variation in a candidate gene prior to performing traditional candidate gene case-control association analysis in an attempt to identify the genetic determinants responsible for late onset sensorineural hearing loss (see section 1.10 for a detailed description of the aims of this project). Late onset hearing loss is a disabling disease exhibited by a considerable proportion of the ageing population and the major cause of this disease appears to be loss of sensory hair cells from the cochlea of the inner ear (see discussion in section 1.5). Brn-3c the candidate gene investigated for this project, a POU domain transcription factor that within the inner ear is specifically expressed in hair cells appears crucial for long-term hair cell survival (Erkman et al, 1996; Xiang et al, 1997; 1998; Vahava et al, 1998; a thorough discussion of the evidence that Brn-3c is a good candidate gene for susceptibility to late onset hearing loss is presented in section 1.8.1). Thus, it was hypothesised that common inter-individual sequence variants in the Brn-3c gene or within the regulatory regions of this gene that have a subtle affect on function or expression of Brn-3c respectively, may affect the genetic risk for late onset hearing loss (the aims of this study are summarised in Chapter 2.0).

At the onset of this project genetic variation in the Brn-3c gene had not been extensively studied. The data presented in this thesis constitutes the first comprehensive report of *common* genetic variation in the Brn-3c gene. Mutation scanning of the Brn-3c gene by PCR-SSCP analysis identified seven sequence variants at the Brn-3c locus five of which are novel and two that were previously reported in the NCBI SNP database but non-validated (see Fig. 4.16, section 4.5, page 162). Of these, two of the novel sequence variants are biallelic single nucleotide substitutions: -386C>A within the 5'-flanking region of the Brn-3c gene and +90C>T within exon one of the gene. However, the +90C>T single nucleotide substitution is a synonymous sequence variant and does not change an amino acid and both of these sequence variants appear rare based on the results of the PCR-SSCP screen on 45 individuals. Therefore, these sequence variants were not pursued further within this project (see section 4.5 page 163 for a greater discussion of the significance of these two sequence variants). Of the remaining five sequence variants identified at the Brn-3c locus all are within the 5'-flanking region of the Brn-3c gene and appear common in the general population (based on the results of the PCR-SSCP screen on 45 individuals). Three are novel multi-allelic repeat sequence variants: -3495$(GT)_{15\text{-}21,24,}$ -3432poly-G polymorphism and -566$(GT)_{17\text{-}23}$ and two: -

3457$(GA)_{1-3}$, a polymorphic dinucleotide repeat and -1391A>C, a single nucleotide substitution were previously reported in the NCBI SNP database but non-validated. Thus, raising the possibility that any or a combination of these common sequence variants may affect expression levels of the Brn-3c gene and hence, modify the genetic risk for late onset hearing loss. Consequently, all were candidates for functional analysis.

By adopting an *in-vitro functional screening approach* of the common sequence variants in the Brn-3c 5'-flanking region utilising comparative EMSA analysis and transient transfection reporter gene assay the data presented in this thesis suggest that at least three common sequence variants in the Brn-3c promoter: -3432poly-G polymorphism, -1391A>C and the -566$(GT)_n$ repeat in the context of the -3432polyG polymorphism could have a functional affect on Brn-3c regulation. It is important to appreciate that the physiological significance of these results is unknown. The inaccessibility of the human inner ear and the lack of human sensory hair cell lines prevent *in-vivo* confirmation of the effect of these sequence variants (for a greater discussion of the limitations of functional data generated *in-vitro* see section 5.8.1). Nevertheless, these results do provide insight into the possible functional effect of common sequence variation in the Brn-3c promoter on Brn-3c regulation and undoubtedly provide candidate functional sequence variants for subsequent association based analysis designed to identify the genetic determinants that underlie late onset hearing loss. *In-vitro* analysis shows that both the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism (section 6.3.1) and the -1391A>C single nucleotide substitution (section 7.4) have a subtle affect on basal activity of the Brn-3c promoter. Thus, it is possible that either alone or in combination these sequence variants could lead to inter-individual variations in Brn-3c expression as small differences in Brn-3c promoter activity such those reported in this thesis may have physiological effects *in-vivo*. Indeed it is tempting to speculate that subtle differences in basal activity of the Brn-3c promoter could have cumulative effects such that hair cell survival is comprised after a certain period of time.

By using a dominant negative approach *in-vitro* analysis suggests that both the -3432poly-G polymorphism (section 5.4.1) and the -566$(GT)_n$ repeat in the context of the -3432poly-G polymorphism (section 6.3.2) can modulate Brn-3c promoter activity under conditions where SP1 is limiting. Specifically, the data suggests a functional interplay between the -566$(GT)_n$ repeat and the SP1 binding site in the -3432poly-G sequence is a determinant of the response of the Brn-3c promoter to SP1. It is difficult

to speculate on the physiological significance of these results. Although, SP1 is an extremely well characterised transcription factor and is implicated in a wide range of cellular responses (for review see Kaczynski et al, 2003) nothing is known of its role in sensory hair cells. Further analysis that is beyond the scope of this PhD project is required to identify the molecular pathways that regulate SP1 activity in hair cells. However, these results do raise the possibility that allele specific effects induced by common repeat sequence variants in the Brn-3c promoter may not have a significant affect on Brn-3c expression by acting in isolation, but rather the combined effect(s) of these repeat variants could be more important. Replication of these findings with the naturally occurring common 5' haplotypes in the Brn-3c promoter would help confirm the significance of these results (Brn-3c promoter-luciferase reporter gene constructs that varied in the length of the GT repeat at -566 were created using controlled haplotypes at -566 in combination with the native haplotypes 5' of the -3432poly-G polymorhism; see section 6.2). Furthermore, the fact the -3432poly-G polymorphism is a second similar polymorphism consisting of a poly-guanine repeat that exhibits multiple single base substitutions that has been shown to affect promoter activity (the poly-G polymorphism at -3432 in the Brn-3c promoter is very similar to one previously identified in the promoter of the KLK1 gene that is associated with hypertensive end stage renal disease; Song et al, 1997; Yu et al, 2002; see also section 5.2) raises the possibility that other poly-guanine repeats of this nature may be polymorphic and has implications for other genes and susceptibility to disease.

The long-term aim of this research project is to use the candidate functional sequence variants in the Brn-3c gene identified from the *in-vitro* functional screen in association based analysis in an attempt to identify the genetic determinants that underlie late onset hearing loss. At the onset of this project well-characterised patient cohorts for late onset hearing loss were scarce. Consequently, a major aim of this PhD project was to initiate collection of a large and well-characterised patient cohort for late onset hearing loss that could be used for a preliminary association study within the time constraints of the PhD project and subsequently, for a much larger population based case-control association study that is beyond the scope of the PhD project. It was accepted at the outset of this project that the number of samples it would be possible to recruit within the time frame of the PhD project would have an impact on the power of the preliminary association study.

Collection of a well-characterised patient cohort for late onset hearing loss was successfully initiated from the adult hearing aid clinic at the Royal Free Nose, Ear and

Throat Hospital, London, U.K. in collaboration with consultant audiologist, Dr. Barbara Cadge (see section 8.2). During the course of this project 142 patients were successfully recruited for the preliminary association study and sample collection is currently continuing with the aim to collect at least 1000 well-characterised samples. Both the -3432poly-G polymorphism (section 8.4) and the -1391A>C single nucleotide substitution (section 8.6) were genotyped in our late onset hearing loss cohort alongside a similar number of general population samples to establish allele frequency and for preliminary association analysis. Subtle differences in allele frequency for both the -3432poly-G polymorphism and the -1391A>C single nucleotide substitution were found between the late onset hearing loss and general population sample groups examined, but were not significantly different when examined by χ^2 analysis (see section 8.5 and 8.7, respectively). However, as discussed previously (section 8.8) this cannot be taken as conclusive evidence that these sequence variants do not modify the genetic risk for late onset hearing loss. Several shortcomings in the design of the preliminary association study coupled with the multi-factorial nature of late onset hearing loss, a complex disease, hinder the likelihood of finding a positive association in a small-scale pilot study such as the preliminary association study presented in this thesis.

In conclusion, further analysis that is beyond the time-frame of this PhD project is required to obtain conclusive evidence whether the -3432poly-G polymorphism or the -1391A>C single nucleotide substitution or both of these sequence variants underlie susceptibility to late onset hearing loss. Continued recruitment of patients with late onset hearing loss into our *London* cohort and novel sample collection from two retrospective *Scottish* cohorts means our laboratory will soon be in a position to undertake association analysis on a large-scale (see section 8.8, page 308). Large-scale case-control association analysis would allow one to determine with certainty whether either of these sequence variants are associated with late onset hearing loss. Alternatively, treating late onset hearing loss as a quantitative trait and using these large sample cohorts to perform quantitative trait analysis may be a more productive approach to find an association (see Fransen et al, 2004; see also section 1.6.3). If a positive association were identified further functional analysis in addition to that presented in this thesis would be required to determine with more certainty whether either or both of these common sequence variants are causative for late onset hearing loss (this is discussed in section 8.9, page 312). For example, it would be especially important to establish if the SP1-mediated differential activity of Brn-3c promoter-luciferase reporter gene constructs that vary in the nature of the -3432poly-G allele is dependent on the

length of the native GT repeat at -3495. As it may well be that the -3432poly-G polymorphism is only functional in the context of its native 5' haplotype. Certainly, the importance of 5'haplotypes on Brn-3c regulation is illustrated by functional analysis incorporating the effect of the -566$(GT)_n$ polymorphism (*in-vitro* evidence indicates that the length of the -566$(GT)_n$ repeat in the context of the -3432 poly-G polymorphism plays an important role in SP1-mediated regulation of Brn-3c; section 6.3.2). Ultimately, it is important to consider that the effect of common sequence variants in the Brn-3c gene on late onset hearing loss disease risk may only manifest when present in combination with variation in another gene at another locus and / or upon repeated exposure to environmental stress. In this regard the future genome wide association studies that are anticipated upon completion of the HapMap when the common patterns of linkage disequilibrium throughout the human genome are established will likely be a valuable tool to determine if haplotypes spanning multi-genes can be identified that underlie susceptibility to late onset hearing loss (for more information on the HapMap project see section 1.6.2.1). However, a shortcoming of the HapMap is that it is focussing on linkage diseqilibrium between common *single nucleotide substitutions* and the affect of common repetitive sequence variants on disease risk may be missed. Inter-individual variations in gene expression due to common *repetitive sequence variants* in the promoters of causative genes may be just as important as the effect of single nucleotide substitutions on complex disease risk. Certainly, with the advent of high through-put genotyping platforms, such as the ABI Prism 3100 genetic analyser (Applied Biosystems) genotyping polymorphic dinucleotide repeats is relatively straight forward by using fluorescently tagged primers in a PCR reaction and then exploiting differences in repeat length. Although, genotyping complex polymorphisms such as the -3432poly-G polymorphism, which vary in mononucleotide repeat length and exhibit single base substitutions within the mononucleotide repeat is currently technically more demanding and time-consuming. Recently it has been reported that mononucleotide repeats are "a neglected polymorphism for generating high density genetic maps *in silico*" (Cohen et al, 2004). Cohen et al, developed a method to rapidly genotype mononucleotide repeats based on sizing of fluorescent-labelled primer extension products. The method was tested and the accuracy confirmed by genotyping 16 mononucleotide repeats, 14 A/T repeats, along human chromosome 22 and comparing the data to that obtained by direct sequencing; 10 mononucleotide repeats were found to be polymorphic. However, such a method, or similar, is unlikely to be suitable for accurate genotyping of mononucleotide G or C repeats. The data presented in this thesis

coupled with that of Song et al, 1997 and Yu et al, 2002 suggests that tracts of poly-G / Poly-C may also exhibit single nucleotide substitutions with the repeat; alleles exhibiting single nucleotide substitutions would be missed in genotyping methods such as that devised by Cohen et al, which exploit differences in length of mononucleotide repeats. Consequently, the value of complex polymorphic poly-G / poly-C repeats in association based studies is currently hindered by a lack of suitable high-throughput genotyping technologies. In addition, the tendency for such polymorphisms to be multi-allelic with a high allele number (Song et al, 1997; Yu et al, 2002; Cohen et al, 2004; data presented in this thesis) means the value of including them in association based studies can probably only be achieved when sufficiently large cohorts are available so as not to reduce the power of the association study. This is despite the fact they may well be functional and have an impact on gene expression.

Currently, association based analysis as a method to dissect the genetic factors that underlie susceptibility to late onset hearing loss or any complex disease is compromised by a lack of knowledge about which sequence variants within the human genome are functional. Synonymous and non-synonymous sequence variants within protein coding regions of a candidate gene, sequence variants that disrupt putative cis-acting elements within the 5'- and 3'- flanking region, and the 5'- and 3'-untranslated region of a candidate gene are *all* putative functional variants. However, in the absence of *in-vitro* and *in-vivo* functional analysis this cannot be confirmed and such putative functional sequence variants cannot be narrowed down to small numbers of candidate variants for testing in association based analysis. Consequently, when designing association-based studies *all* such putative functional variants should be considered as causative. Van Laer et al, 2006 have recently reported an association between sequence variants in KCNE1, KCNQ1 and KCNQ4 and susceptibility to NIHL. Moreover, based on patch-clamp experiments they propose p.D85N a non-synonymous variant in the potassium channel subunit gene, KCNE1, as a possible causative variant for NIHL. This study was based on performing case-control association analysis on each of 35 single nucleotide substitutions found within the genomic region of 10 candidate genes involved in cell-cell communication and potassium homeostasis in the inner ear. However, only p.D85N was selected for subsequent function analysis. This is despite finding several other position associations. Hence, although this study supports the value of association based analysis to delineate the genetic determinants that underlie late onset hearing loss; in the absence of thorough functional screening it is by no means conclusive. Without undertaking a functional screen for all variants for which an

association was identified one cannot conclude with certainty that p.D85N is causative. In addition, the 5' and 3' regulatory regions of the candidate genes selected for inclusion in this study was overlooked; p.D85N could be in linkage with a functionally significant variant in the KCNE1 promoter. Thus, the value of the functional screening approach presented in this thesis is undeniable. By obtaining functional knowledge of *all* the common sequence variants within the coding and regulatory regions of a candidate gene more informative conclusions from association-based studies will be able to be drawn.

To conclude, by identifying the genetic determinants that underlie susceptibility to late onset hearing loss the underlying pathophysiology of this highly complex and disabling disease will begin to be unravelled by delineation of the molecular pathways and key components involved. This would be highly significant, as it would provide a platform for future therapeutic intervention to ultimately halt loss of hearing with increasing age and not least, it would aid our understanding of the hearing process by building upon insight gained from the study of rare monogenic congenital forms of deafness.

Scientific corrections for this thesis are on page 367.

References

Ackerman,S.L., Minden,A.G., Williams,G.T., Bobonis,C., and Yeung,C.Y. 1991. Functional significance of an overlapping consensus binding motif for Sp1 and Zif268 in the murine adenosine deaminase gene promoter. Proc. Natl. Acad. Sci. U. S. A 88: 7523-7527.

Ahituv,N. and Avraham,K.B. 2002. Mouse models for human deafness: current tools for new fashions. Trends Mol Med. 8: 447-451.

Ahmed,Z.M., Riazuddin,S., Riazuddin,S., and Wilcox,E.R. 2003. The molecular genetics of Usher syndrome. Clin. Genet. 63: 431-444.

Akai,J., Kimura,A., and Hata,R.I. 1999. Transcriptional regulation of the human type I collagen alpha2 (COL1A2) gene by the combination of two dinucleotide repeats. Gene 239: 65-73.

Al Sarraj,A., Day,R.M., and Thiel,G. 2005. Specificity of transcriptional regulation by the zinc finger transcription factors Sp1, Sp3, and Egr-1. J. Cell Biochem. 94: 153-167.

Alam,S.A., Oshima,T., Suzuki,M., Kawase,T., Takasaka,T., and Ikeda,K. 2001. The expression of apoptosis-related proteins in the aged cochlea of Mongolian gerbils. Laryngoscope 111: 528-534.

Albanese,V., Biguet,N.F., Kiefer,H., Bayard,E., Mallet,J., and Meloni,R. 2001. Quantitative effects on gene silencing by allelic variation at a tetranucleotide microsatellite. Hum. Mol Genet. 10: 1785-1792.

Alkalay,I., Yaron,A., Hatzubai,A., Jung,S., Avraham,A., Gerlitz,O., Pashut-Lavon,I., and Ben Neriah,Y. 1995. In vivo stimulation of I kappa B phosphorylation is not sufficient to activate NF-kappa B. Mol Cell Biol. 15: 1294-1301.

Altmuller,J., Palmer,L.J., Fischer,G., Scherb,H., and Wjst,M. 2001. Genomewide scans of complex human diseases: true linkage is hard to find. Am. J. Hum. Genet. 69: 936-950.

Altshuler,D., Hirschhorn,J.N., Klannemark,M., Lindgren,C.M., Vohl,M.C., Nemesh,J., Lane,C.R., Schaffner,S.F., Bolk,S., Brewer,C., Tuomi,T., Gaudet,D., Hudson,T.J., Daly,M., Groop,L., and Lander,E.S. 2000. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat. Genet. 26: 76-80.

Ammanamanchi,S. and Brattain,M.G. 2001. Sp3 is a transcriptional repressor of transforming growth factor-beta receptors. J. Biol. Chem. 276: 3348-3352.

Andersen,B. and Rosenfeld,M.G. 2001. POU domain factors in the neuroendocrine system: lessons from developmental biology provide insights into human disease. Endocr. Rev. 22: 2-35.

Aruga,J. 2004. The role of Zic genes in neural development. Mol Cell Neurosci. 26: 205-221.

Ashmore,J.F., Geleoc,G.S., and Harbott,L. 2000. Molecular mechanisms of sound amplification in the mammalian cochlea. Proc. Natl. Acad. Sci. U. S. A 97: 11759-11764.

Ashmore,J.F. and Mammano,F. 2001. Can you still see the cochlea for the molecules? Curr. Opin. Neurobiol. 11: 449-454.

Aurora,R. and Herr,W. 1992. Segments of the POU domain influence one another's DNA-binding specificity. Mol Cell Biol. 12: 455-467.

Avraham,K.B. 2003. Mouse models for deafness: lessons for the human inner ear and hearing loss. Ear Hear. 24: 332-341.

Baba,S., Kukita,Y., Higasa,K., Tahira,T., and Hayashi,K. 2003. Single-stranded conformational polymorphism analysis using automated capillary array electrophoresis apparatuses. Biotechniques 34: 746-750.

Barald,K.F. and Kelley,M.W. 2004. From placode to polarization: new tunes in inner ear development. Development 131: 4119-4130.

Bateman,E. 1998. Autoregulation of eukaryotic transcription factors. Prog. Nucleic Acid Res. Mol Biol. 60: 133-168.

Beisel,K.W., Nelson,N.C., Delimont,D.C., and Fritzsch,B. 2000. Longitudinal gradients of KCNQ4 expression in spiral ganglion and cochlear hair cells correlate with progressive hearing loss in DFNA2. Brain Res. Mol Brain Res. 82: 137-149.

Belyantseva,I.A., Adler,H.J., Curi,R., Frolenkov,G.I., and Kachar,B. 2000. Expression and localization of prestin and the sugar transporter GLUT-5 during development of electromotility in cochlear outer hair cells. J. Neurosci. 20: RC116.

Bermingham-McDonogh,O. and Rubel,E.W. 2003. Hair cell regeneration: winging our way towards a sound future. Curr. Opin. Neurobiol. 13: 119-126.

Bermingham,N.A., Hassan,B.A., Price,S.D., Vollrath,M.A., Ben Arie,N., Eatock,R.A., Bellen,H.J., Lysakowski,A., and Zoghbi,H.Y. 1999. Math1: an essential gene for the generation of inner ear hair cells. Science 284: 1837-1841.

Bertina,R.M., Koeleman,B.P., Koster,T., Rosendaal,F.R., Dirven,R.J., de Ronde,H., van der Velden,P.A., and Reitsma,P.H. 1994. Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 369: 64-67.

Bichara,M., Pinet,I., Schumacher,S., and Fuchs,R.P. 2000. Mechanisms of dinucleotide repeat instability in Escherichia coli. Genetics 154: 533-542.

Biggin,M.D. 2001. To bind or not to bind. Nat. Genet. 28: 303-304.

Bitner-Glindzicz,M., Lindley,K.J., Rutland,P., Blaydon,D., Smith,V.V., Milla,P.J., Hussain,K., Furth-Lavi,J., Cosgrove,K.E., Shepherd,R.M., Barnes,P.D., O'Brien,R.E., Farndon,P.A., Sowden,J., Liu,X.Z., Scanlan,M.J., Malcolm,S., Dunne,M.J., Aynsley-Green,A., and Glaser,B. 2000. A recessive contiguous gene deletion causing infantile hyperinsulinism, enteropathy and deafness identifies the Usher type 1C gene. Nat. Genet. 26: 56-60.

Bitner-Glindzicz,M. 2002. Hereditary deafness and phenotyping in humans. Br. Med. Bull. 63: 73-94.

Blackwood,E.M. and Kadonaga,J.T. 1998. Going the distance: a current view of enhancer action. Science 281: 60-63.

Boeda,B., El Amraoui,A., Bahloul,A., Goodyear,R., Daviet,L., Blanchard,S., Perfettini,I., Fath,K.R., Shorte,S., Reiners,J., Houdusse,A., Legrain,P., Wolfrum,U., Richardson,G., and Petit,C. 2002. Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle. EMBO J. 21: 6689-6699.

Bohnker,B.K., Page,J.C., Rovig,G., Betts,L.S., Muller,J.G., and Sack,D.M. 2002. U.S. Navy and Marine Corps Hearing Conservation Program, 1995-1999: mean hearing thresholds for enlisted personnel by gender and age groups. Mil. Med. 167: 132-135.

Bolz,H., von Brederlow,B., Ramirez,A., Bryda,E.C., Kutsche,K., Nothwang,H.G., Seeliger,M., del,C.-S., Vila,M.C., Molina,O.P., Gal,A., and Kubisch,C. 2001. Mutation of CDH23, encoding a new member of the cadherin gene family, causes Usher syndrome type 1D. Nat. Genet. 27: 108-112.

Bork,J.M., Peters,L.M., Riazuddin,S., Bernstein,S.L., Ahmed,Z.M., Ness,S.L., Polomeno,R., Ramesh,A., Schloss,M., Srisailpathy,C.R., Wayne,S., Bellman,S., Desmukh,D., Ahmed,Z., Khan,S.N., Kaloustian,V.M., Li,X.C., Lalwani,A., Riazuddin,S., Bitner-Glindzicz,M., Nance,W.E., Liu,X.Z., Wistow,G., Smith,R.J., Griffith,A.J., Wilcox,E.R., Friedman,T.B., and Morell,R.J. 2001. Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFNB12 are caused by allelic mutations of the novel cadherin-like gene CDH23. Am. J. Hum. Genet. 68: 26-37.

Borrmann,L., Seebeck,B., Rogalla,P., and Bullerdiek,J. 2003. Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)-repeat. Oncogene 22: 756-760.

Botstein,D. and Risch,N. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33 Suppl: 228-237.

Boyer,J.C., Yamada,N.A., Roques,C.N., Hatch,S.B., Riess,K., and Farber,R.A. 2002. Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. Hum. Mol Genet. 11: 707-713.

Brooker,R., Hozumi,K., and Lewis,J. 2006. Notch ligands with contrasting functions: Jagged1 and Delta1 in the mouse inner ear. Development 133: 1277-1286.

Bryant,J., Goodyear,R.J., and Richardson,G.P. 2002. Sensory organ development in the inner ear: molecular and cellular mechanisms. Br. Med. Bull. 63: 39-57.

Bryda,E.C., Ling,H., and Flaherty,L. 1997. A high-resolution genetic map around waltzer on mouse chromosome 10 and identification of a new allele of waltzer. Mamm. Genome 8: 1-4.

Bryda,E.C., Kim,H.J., Legare,M.E., Frankel,W.N., and Noben-Trauth,K. 2001. High-resolution genetic and physical mapping of modifier-of-deafwaddler (mdfw) and Waltzer (Cdh23v). Genomics 73: 338-342.

Budhram-Mahadeo,V., Morris,P.J., Lakin,N.D., Theil,T., Ching,G.Y., Lillycrop,K.A., Moroy,T., Liem,R.K., and Latchman,D.S. 1995. Activation of the alpha-internexin promoter by the Brn-3a transcription factor is dependent on the N-terminal region of the protein. J. Biol. Chem. 270: 2853-2858.

Budhram-Mahadeo,V., Parker,M., and Latchman,D.S. 1998. POU transcription factors Brn-3a and Brn-3b interact with the estrogen receptor and differentially regulate transcriptional activity via an estrogen response element. Mol Cell Biol. 18: 1029-1041.

Bulfone,A., Menguzzato,E., Broccoli,V., Marchitiello,A., Gattuso,C., Mariani,M., Consalez,G.G., Martinez,S., Ballabio,A., and Banfi,S. 2000. Barhl1, a gene belonging to a new subfamily of mammalian homeobox genes, is expressed in migrating neurons of the CNS. Hum. Mol Genet. 9: 1443-1452.

Cai,Q.Q. and Touitou,I. 1993. Excess PCR primers may dramatically affect SSCP efficiency. Nucleic Acids Res. 21: 3909-3910.

Cambien,F., Poirier,O., Lecerf,L., Evans,A., Cambou,J.P., Arveiler,D., Luc,G., Bard,J.M., Bara,L., Ricard,S., and . 1992. Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. Nature 359: 641-644.

Campbell,D.J. 2001. The kallikrein-kinin system in humans. Clin. Exp. Pharmacol Physiol 28: 1060-1065.

Campbell,K. 2003. Ototoxicity: understanding oxidative mechanisms. J. Am. Acad. Audiol. 14: 121-123.

Cardon,L.R. and Bell,J.I. 2001. Association study designs for complex diseases. Nat. Rev. Genet. 2: 91-99.

Cardon,L.R. and Abecasis,G.R. 2003. Using haplotype blocks to map human complex trait loci. Trends Genet. 19: 135-140.

Cartharius,K., Frech,K., Grote,K., Klocke,B., Haltmeier,M., Klingenhoff,A., Frisch,M., Bayerlein,M., and Werner,T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 21: 2933-2942.

Castellvi-Bel,S., Sanchez,A., Badenas,C., Mallolas,J., Barcelo,A., Jimenez,D., Villa,M., Estivill,X., and Mila,M. 1999. Single-strand conformation polymorphism analysis in the FMR1 gene. Am. J. Med. Genet. 84: 262-265.

Catlett-Falcone,R., Landowski,T.H., Oshiro,M.M., Turkson,J., Levitzki,A., Savino,R., Ciliberto,G., Moscinski,L., Fernandez-Luna,J.L., Nunez,G., Dalton,W.S., and Jove,R. 1999. Constitutive activation of Stat3 signaling confers resistance to apoptosis in human U266 myeloma cells. Immunity. 10: 105-115.

Chan,S.L. and Yu,V.C. 2004. Proteins of the bcl-2 family in apoptosis signalling: from mechanistic insights to therapeutic opportunities. Clin. Exp. Pharmacol Physiol 31: 119-128.

Chaurand,P., Luetzenkirchen,F., and Spengler,B. 1999. Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. J. Am. Soc. Mass Spectrom. 10: 91-103.

Chen,G.D. and Fechter,L.D. 2003. The relationship between noise-induced hearing loss and hair cell loss in rats. Hear. Res. 177: 81-90.

Chen,P., Johnson,J.E., Zoghbi,H.Y., and Segil,N. 2002. The role of Math1 in inner ear development: Uncoupling the establishment of the sensory primordium from hair cell fate determination. Development 129: 2495-2505.

Cheng,A.G., Cunningham,L.L., and Rubel,E.W. 2005. Mechanisms of hair cell death and protection. Curr. Opin. Otolaryngol. Head Neck Surg. 13: 343-348.

Clarke,L.A., Rebelo,C.S., Goncalves,J., Boavida,M.G., and Jordan,P. 2001. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol Pathol. 54: 351-353.

Cline,J., Braman,J.C., and Hogrefe,H.H. 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. 24: 3546-3551.

Clough,R.L., Sud,R., Davis-Silberman,N., Hertzano,R., Avraham,K.B., Holley,M., and Dawson,S.J. 2004. Brn-3c (POU4F3) regulates BDNF and NT-3 promoter activity. Biochem. Biophys. Res. Commun. 324: 372-381.

Cohen,H., Danin-Poleg,Y., Cohen,C.J., Sprecher,E., Darvasi,A., and Kashi,Y. 2004. Mono-nucleotide repeats (MNRs): a neglected polymorphism for generating high density genetic maps in silico. Hum. Genet. 115: 213-220.

Collins,J.S. and Schwartz,C.E. 2002. Detecting polymorphisms and mutations in candidate genes. Am. J. Hum. Genet. 71: 1251-1252.

Conne,B., Stutz,A., and Vassalli,J.D. 2000. The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? Nat. Med. 6: 637-641.

Contente,A., Dittmer,A., Koch,M.C., Roth,J., and Dobbelstein,M. 2002. A polymorphic microsatellite that mediates induction of PIG3 by p53. Nat. Genet. 30: 315-320.

Cook,T., Gebelein,B., and Urrutia,R. 1999. Sp1 and its likes: biochemical and functional predictions for a growing family of zinc finger transcription factors. Ann. N. Y. Acad. Sci. 880: 94-102.

Corder,E.H., Saunders,A.M., Strittmatter,W.J., Schmechel,D.E., Gaskell,P.C., Small,G.W., Roses,A.D., Haines,J.L., and Pericak-Vance,M.A. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261: 921-923.

Corey,D.P. and Hudspeth,A.J. 1983. Kinetics of the receptor current in bullfrog saccular hair cells. J. Neurosci. 3: 962-976.

Corey,D.P., Garcia-Anoveros,J., Holt,J.R., Kwan,K.Y., Lin,S.Y., Vollrath,M.A., Amalfitano,A., Cheung,E.L., Derfler,B.H., Duggan,A., Geleoc,G.S., Gray,P.A., Hoffman,M.P., Rehm,H.L., Tamasauskas,D., and Zhang,D.S. 2004. TRPA1 is a candidate for the mechanosensitive transduction channel of vertebrate hair cells. Nature 432: 723-730.

Cotton,R.G., Rodrigues,N.R., and Campbell,R.D. 1988. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc. Natl. Acad. Sci. U. S. A 85: 4397-4401.

Cotton,R.G. 1993. Current methods of mutation detection. Mutat. Res. 285: 125-144.

Cotton,R.G. and Scriver,C.R. 1998. Proof of "disease causing" mutation. Hum. Mutat. 12: 1-3.

Coucke,P.J., Van Hauwe,P., Kelley,P.M., Kunst,H., Schatteman,I., Van Velzen,D., Meyers,J., Ensink,R.J., Verstreken,M., Declau,F., Marres,H., Kastury,K., Bhasin,S., McGuirt,W.T., Smith,R.J., Cremers,C.W., Van de,H.P., Willems,P.J., Smith,S.D., and Van Camp,G. 1999. Mutations in the KCNQ4 gene are responsible for autosomal dominant deafness in four DFNA2 families. Hum. Mol Genet. 8: 1321-1328.

Courey,A.J. and Tjian,R. 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55: 887-898.

Crawley,E., Kay,R., Sillibourne,J., Patel,P., Hutchinson,I., and Woo,P. 1999. Polymorphic haplotypes of the interleukin-10 5' flanking region determine variable interleukin-10 transcription and are associated with particular phenotypes of juvenile rheumatoid arthritis. Arthritis Rheum. 42: 1101-1108.

Crosby,S.D., Puetz,J.J., Simburger,K.S., Fahrner,T.J., and Milbrandt,J. 1991. The early response gene NGFI-C encodes a zinc finger transcriptional activator and is a member of the GCGGGGGCG (GSG) element-binding protein family. Mol Cell Biol. 11: 3835-3841.

Dachet,C., Poirier,O., Cambien,F., Chapman,J., and Rouis,M. 2000. New functional promoter polymorphism, CETP/-629, in cholesteryl ester transfer protein (CETP) gene related to CETP mass and high density lipoprotein cholesterol levels: role of Sp1/Sp3 in transcriptional regulation. Arterioscler. Thromb. Vasc. Biol. 20: 507-515.

Dahlback,B. 1997. Resistance to activated protein C caused by the factor VR506Q mutation is a common risk factor for venous thrombosis. Thromb. Haemost. 78: 483-488.

Dahlman,I., Eaves,I.A., Kosoy,R., Morrison,V.A., Heward,J., Gough,S.C., Allahabadia,A., Franklyn,J.A., Tuomilehto,J., Tuomilehto-Wolf,E., Cucca,F., Guja,C., Ionescu-Tirgoviste,C., Stevens,H., Carr,P., Nutland,S., McKinney,P., Shield,J.P., Wang,W., Cordell,H.J., Walker,N., Todd,J.A., and Concannon,P. 2002. Parameters for reliable results in genetic association studies in common disease. Nat. Genet. 30: 149-150.

Daley,G.Q. and Cargill,M. 2001. The heart SNPs a beat: polymorphisms in candidate genes for cardiovascular disease. Trends Cardiovasc. Med. 11: 60-66.

Dallos,P. and Fakler,B. 2002. Prestin, a new type of motor protein. Nat. Rev. Mol Cell Biol. 3: 104-111.

Das,P.M., Ramachandran,K., vanWert,J., and Singal,R. 2004. Chromatin immunoprecipitation assay. Biotechniques 37: 961-969.

Daudet,N. and Lewis,J. 2005. Two contrasting roles for Notch activity in chick inner ear development: specification of prosensory patches and lateral inhibition of hair-cell differentiation. Development 132: 541-551.

Davis,A.C. 1989. The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. Int. J. Epidemiol. 18: 911-917.

Davis,A.C. 1995. Hearing in Adults. London: Whurr Publishers Ltd.

Davis,R.R., Newlander,J.K., Ling,X., Cortopassi,G.A., Krieg,E.F., and Erway,L.C. 2001. Genetic basis for susceptibility to noise-induced hearing loss in mice. Hear. Res. 155: 82-90.

Davis,R.R., Kozel,P., and Erway,L.C. 2003. Genetic influences in individual susceptibility to noise: a review. Noise. Health 5: 19-28.

Dawson,S., Hamsten,A., Wiman,B., Henney,A., and Humphries,S. 1991. Genetic variation at the plasminogen activator inhibitor-1 locus is associated with altered levels of plasma plasminogen activator inhibitor-1 activity. Arterioscler. Thromb. 11: 183-190.

Dawson,S.J., Wiman,B., Hamsten,A., Green,F., Humphries,S., and Henney,A.M. 1993. The two allele sequences of a common polymorphism in the promoter of the plasminogen activator inhibitor-1 (PAI-1) gene respond differently to interleukin-1 in HepG2 cells. J. Biol. Chem. 268: 10739-10745.

Dawson,S.J., Liu,Y.Z., Rodel,B., Moroy,T., and Latchman,D.S. 1996a. The ability of POU family transcription factors to activate or repress gene expression is dependent on the spacing and context of their specific response elements. Biochem. J. 314 (Pt 2): 439-443.

Dawson,S.J., Morris,P.J., and Latchman,D.S. 1996b. A single amino acid change converts an inhibitory transcription factor into an activator. J. Biol. Chem. 271: 11631-11633.

Dawson,S.J., Palmer,R.D., Morris,P.J., and Latchman,D.S. 1998. Functional role of position 22 in the homeodomain of Brn-3 transcription factors. Neuroreport 9: 2305-2309.

Day,I.N., Humphries,S.E., Richards,S., Norton,D., and Reid,M. 1995. High-throughput genotyping using horizontal polyacrylamide gels with wells arranged for microplate array diagonal gel electrophoresis (MADGE). Biotechniques 19: 830-835.

de Wet,J.R., Wood,K.V., DeLuca,M., Helinski,D.R., and Subramani,S. 1987. Firefly luciferase gene: structure and expression in mammalian cells. Mol Cell Biol. 7: 725-737.

Deeb,S.S., Fajas,L., Nemoto,M., Pihlajamaki,J., Mykkanen,L., Kuusisto,J., Laakso,M., Fujimoto,W., and Auwerx,J. 1998. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nat. Genet. 20: 284-287.

Delhase,M., Castrillo,J.L., de la,H.M., Rajas,F., and Hooghe-Peters,E.L. 1996. AP-1 and Oct-1 transcription factors down-regulate the expression of the human PIT1/GHF1 gene. J. Biol. Chem. 271: 32349-32358.

Dellas,C. and Loskutoff,D.J. 2005. Historical analysis of PAI-1 from its discovery to its potential role in cell motility and disease. Thromb. Haemost. 93: 631-640.

Demczuk,S., Harbers,M., and Vennstrom,B. 1993. Identification and analysis of all components of a gel retardation assay by combination with immunoblotting. Proc. Natl. Acad. Sci. U. S. A 90: 2574-2578.

Dennig,J., Beato,M., and Suske,G. 1996. An inhibitor domain in Sp3 regulates its glutamine-rich activation domains. EMBO J. 15: 5659-5667.

DeStefano,A.L., Gates,G.A., Heard-Costa,N., Myers,R.H., and Baldwin,C.T. 2003. Genomewide linkage analysis to presbycusis in the Framingham Heart Study. Arch. Otolaryngol. Head Neck Surg. 129: 285-289.

Di Palma,F., Holme,R.H., Bryda,E.C., Belyantseva,I.A., Pellegrino,R., Kachar,B., Steel,K.P., and Noben-Trauth,K. 2001a. Mutations in Cdh23, encoding a new type of cadherin, cause stereocilia disorganization in waltzer, the mouse model for Usher syndrome type 1D. Nat. Genet. 27: 103-107.

Di Palma,F., Pellegrino,R., and Noben-Trauth,K. 2001b. Genomic structure, alternative splice forms and normal and mutant alleles of cadherin 23 (Cdh23). Gene 281: 31-41.

Donahue,M.P. and Allen,A.S. 2005. Genetic association studies in cardiology. Am. Heart J. 149: 964-970.

Donohoe,E. 2004. Denaturing High-Performance Liquid Chromatography Using the WAVE DNA Fragment Analysis System. Methods Mol Med. 108: 173-188.

Douglas,J.A., Erdos,M.R., Watanabe,R.M., Braun,A., Johnston,C.L., Oeth,P., Mohlke,K.L., Valle,T.T., Ehnholm,C., Buchanan,T.A., Bergman,R.N., Collins,F.S., Boehnke,M., and Tuomilehto,J. 2001. The peroxisome proliferator-activated receptor-gamma2 Pro12A1a variant: association with type 2 diabetes and trait differences. Diabetes 50: 886-890.

Duan,M.L., Ulfendahl,M., Laurell,G., Counter,A.S., Pyykko,I., Borg,E., and Rosenhall,U. 2002. Protection and treatment of sensorineural hearing disorders caused by exogenous factors: experimental findings and potential clinical application. Hear. Res. 169: 169-178.

Duan,Z. and Horwitz,M. 2003. Targets of the transcriptional repressor oncoprotein Gfi-1. Proc. Natl. Acad. Sci. U. S. A 100: 5932-5937.

Dumont,R.A., Lins,U., Filoteo,A.G., Penniston,J.T., Kachar,B., and Gillespie,P.G. 2001. Plasma membrane Ca2+-ATPase isoform 2a is the PMCA of hair bundles. J. Neurosci. 21: 5066-5078.

Dynan,W.S. and Tjian,R. 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35: 79-87.

Ebert,S.N. and Wong,D.L. 1995. Differential activation of the rat phenylethanolamine N-methyltransferase gene by Sp1 and Egr-1. J. Biol. Chem. 270: 17299-17305.

El Amraoui,A. and Petit,C. 2005. Usher I syndrome: unravelling the mechanisms that underlie the cohesion of the growing hair bundle in inner ear sensory cells. J. Cell Sci. 118: 4593-4603.

Ellegren,H. 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5: 435-445.

Ellis,T.P., Humphrey,K.E., Smith,M.J., and Cotton,R.G. 1998. Chemical cleavage of mismatch: a new look at an established method. Hum. Mutat. 11: 345-353.

Emmerich,E., Richter,F., Reinhold,U., Linss,V., and Linss,W. 2000. Effects of industrial noise exposure on distortion product otoacoustic emissions (DPOAEs) and hair cell loss of the cochlea--long term experiments in awake guinea pigs. Hear. Res. 148: 9-17.

Eng,S.R., Gratwick,K., Rhee,J.M., Fedtsova,N., Gan,L., and Turner,E.E. 2001. Defects in sensory axon growth precede neuronal death in Brn3a-deficient mice. J. Neurosci. 21: 541-549.

Erkman,L., McEvilly,R.J., Luo,L., Ryan,A.K., Hooshmand,F., O'Connell,S.M., Keithley,E.M., Rapaport,D.H., Ryan,A.F., and Rosenfeld,M.G. 1996. Role of transcription factors Brn-3.1 and Brn-3.2 in auditory and visual system development. Nature 381: 603-606.

Erway,L.C., Willott,J.F., Archer,J.R., and Harrison,D.E. 1993. Genetics of age-related hearing loss in mice: I. Inbred and F1 hybrid strains. Hear. Res. 65: 125-132.

Erway,L.C., Shiau,Y.W., Davis,R.R., and Krieg,E.F. 1996. Genetics of age-related hearing loss in mice. III. Susceptibility of inbred and F1 hybrid strains to noise-induced hearing loss. Hear. Res. 93: 181-187.

Evans,P. and Halliwell,B. 1999. Free radicals and hearing. Cause, consequence, and criteria. Ann. N. Y. Acad. Sci. 884: 19-40.

Farrall,M. and Morris,A.P. 2005. Gearing up for genome-wide gene-association studies. Hum. Mol Genet. 14 Spec No. 2: R157-R162.

Faulkes,D.J., Ensor,E., Le Rouzic,E., and Latchman,D.S. 2004. Distinct domains of Brn-3a regulate apoptosis and neurite outgrowth in vivo. Neuroreport 15: 1421-1425.

Feghali,J.G., Lefebvre,P.P., Staecker,H., Kopke,R., Frenz,D.A., Malgrange,B., Liu,W., Moonen,G., Ruben,R.J., and Van De Water,T.R. 1998. Mammalian auditory hair cell regeneration/repair and protection: a review and future directions. Ear Nose Throat J. 77: 276, 280, 282-276, 280, 285.

Fekete,D.M. and Wu,D.K. 2002. Revisiting cell fate specification in the inner ear. Curr. Opin. Neurobiol. 12: 35-42.

Finney,M., Ruvkun,G., and Horvitz,H.R. 1988. The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. Cell 55: 757-769.

Fischel-Ghodsian,N. 2003. Mitochondrial deafness. Ear Hear. 24: 303-313.

Fischel-Ghodsian,N., Kopke,R.D., and Ge,X. 2004. Mitochondrial dysfunction in hearing loss. Mitochondrion. 4: 675-694.

Forge,A. and Wright,T. 2002. The molecular architecture of the inner ear. Br. Med. Bull. 63: 5-24.

Forsberg,L., de Faire,U., and Morgenstern,R. 2001. Oxidative stress, human genetic variation, and disease. Arch. Biochem. Biophys. 389: 84-93.

Fortnum,H.M., Summerfield,A.Q., Marshall,D.H., Davis,A.C., and Bamford,J.M. 2001. Prevalence of permanent childhood hearing impairment in the United Kingdom and implications for universal neonatal hearing screening: questionnaire based ascertainment study. BMJ 323: 536-540.

Fortunato,G., Marciano,E., Zarrilli,F., Mazzaccara,C., Intrieri,M., Calcagno,G., Vitale,D.F., La Manna,P., Saulino,C., Marcelli,V., and Sacchetti,L. 2004. Paraoxonase and superoxide dismutase gene polymorphisms and noise-induced hearing loss. Clin. Chem. 50: 2012-2018.

Franca,L.T., Carrilho,E., and Kist,T.B. 2002. A review of DNA sequencing techniques. Q. Rev. Biophys. 35: 169-200.

Fransen,E., Lemkens,N., Van Laer,L., and Van Camp,G. 2003. Age-related hearing impairment (ARHI): environmental risk factors and genetic prospects. Exp. Gerontol. 38: 353-359.

Fransen,E., Van Laer,L., Lemkens,N., Caethoven,G., Flothmann,K., Govaerts,P., Van de,H.P., and Van Camp,G. 2004. A novel Z-score-based method to analyze candidate genes for age-related hearing impairment. Ear Hear. 25: 133-141.

Frass,B., Vassen,L., and Moroy,T. 2002. Gene expression of the POU factor Brn-3a is regulated by two different promoters. Biochim. Biophys. Acta 1579: 207-213.

Freund,A.M., Bichara,M., and Fuchs,R.P. 1989. Z-DNA-forming sequences are spontaneous deletion hot spots. Proc. Natl. Acad. Sci. U. S. A 86: 7465-7469.

Fridberger,A., Flock,A., Ulfendahl,M., and Flock,B. 1998. Acoustic overstimulation increases outer hair cell Ca2+ concentrations and causes dynamic contractions of the hearing organ. Proc. Natl. Acad. Sci. U. S. A 95: 7127-7132.

Fried,M. and Crothers,D.M. 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. 9: 6505-6525.

Friedman,T.B., Schultz,J.M., Ben Yosef,T., Pryor,S.P., Lagziel,A., Fisher,R.A., Wilcox,E.R., Riazuddin,S., Ahmed,Z.M., Belyantseva And IA, and Griffith,A.J. 2003. Recent advances in the understanding of syndromic forms of hearing loss. Ear Hear. 24: 289-302.

Frisch,R., Singleton,K.R., Moses,P.A., Gonzalez,I.L., Carango,P., Marks,H.G., and Funanage,V.L. 2001. Effect of triplet repeat expansion on chromatin structure and expression of DMPK and neighboring genes, SIX5 and DMWD, in myotonic dystrophy. Mol Genet. Metab 74: 281-291.

Fritzsch,B. and Beisel,K.W. 2001. Evolution and development of the vertebrate ear. Brain Res. Bull. 55: 711-721.

Fukada,T., Hibi,M., Yamanaka,Y., Takahashi-Tezuka,M., Fujitani,Y., Yamaguchi,T., Nakajima,K., and Hirano,T. 1996. Two signals are necessary for cell proliferation induced by a cytokine receptor gp130: involvement of STAT3 in anti-apoptosis. Immunity. 5: 449-460.

Gan,L., Xiang,M., Zhou,L., Wagner,D.S., Klein,W.H., and Nathans,J. 1996. POU domain factor Brn-3b is required for the development of a large set of retinal ganglion cells. Proc. Natl. Acad. Sci. U. S. A 93: 3920-3925.

Gan,L., Wang,S.W., Huang,Z., and Klein,W.H. 1999. POU domain factor Brn-3b is essential for retinal ganglion cell differentiation and survival but not for initial cell fate specification. Dev. Biol. 210: 469-480.

Gardemann,A., Lohre,J., Katz,N., Tillmanns,H., Hehrlein,F.W., and Haberbosch,W. 1999. The 4G4G genotype of the plasminogen activator inhibitor 4G/5G gene polymorphism is associated with coronary atherosclerosis in patients at high risk for this disease. Thromb. Haemost. 82: 1121-1126.

Garner,M.M. and Revzin,A. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res. 9: 3047-3060.

Gates,G.A., Couropmitree,N.N., and Myers,R.H. 1999. Genetic associations in age-related hearing thresholds. Arch. Otolaryngol. Head Neck Surg. 125: 654-659.

Gates,G.A. and Mills,J.H. 2005. Presbycusis. Lancet 366: 1111-1120.

Gebhardt,F., Zanker,K.S., and Brandt,B. 1999. Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J. Biol. Chem. 274: 13176-13180.

Gerido,D.A. and White,T.W. 2004. Connexin disorders of the ear, skin, and lens. Biochim. Biophys. Acta 1662: 159-170.

Germer,S. and Higuchi,R. 1999. Single-tube genotyping without oligonucleotide probes. Genome Res. 9: 72-78.

Gerrard.B. and Dean.M. 1998. Single-stranded conformation polymorphism and heteroduplex analysis. In: Cotton.R.G.H., Edkins.E. and Forrest.S., Editors. Mutation Detection: a practical approach. New York: Oxford University Press. P25-34.

Gerrero,M.R., McEvilly,R.J., Turner,E., Lin,C.R., O'Connell,S., Jenne,K.J., Hobbs,M.V., and Rosenfeld,M.G. 1993. Brn-3.0: a POU-domain protein expressed in the sensory, immune, and endocrine systems that functions on elements distinct from known octamer motifs. Proc. Natl. Acad. Sci. U. S. A 90: 10841-10845.

Gidoni,D., Dynan,W.S., and Tjian,R. 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature 312: 409-413.

Gilks,C.B., Bear,S.E., Grimes,H.L., and Tsichlis,P.N. 1993. Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma lines to IL-2-independent growth following activation of a gene (Gfi-1) encoding a novel zinc finger protein. Mol Cell Biol. 13: 1759-1768.

Gill,G., Pascal,E., Tseng,Z.H., and Tjian,R. 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. Proc. Natl. Acad. Sci. U. S. A 91: 192-196.

Gillespie,P.G. and Walker,R.G. 2001. Molecular basis of mechanosensory transduction. Nature 413: 194-202.

Glavac,D. and Dean,M. 1993. Optimization of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. Hum. Mutat. 2: 404-414.

Gong,T.W., Hegeman,A.D., Shin,J.J., Adler,H.J., Raphael,Y., and Lomax,M.I. 1996. Identification of genes expressed after noise exposure in the chick basilar papilla. Hear. Res. 96: 20-32.

Gorman,C.M. 1985. High efficiency gene transfer into mammalian cells. Oxford IRL Press.

Gragg,H., Harfe,B.D., and Jinks-Robertson,S. 2002. Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in Saccharomyces cerevisiae. Mol Cell Biol. 22: 8756-8762.

Gratton,M.A. and Vazquez,A.E. 2003. Age-related hearing loss: current research. Curr. Opin. Otolaryngol. Head Neck Surg. 11: 367-371.

Green,F.R. 2001. Fibrinogen polymorphisms and atherothrombotic disease. Ann. N. Y. Acad. Sci. 936: 549-559.

Grimes,H.L., Chan,T.O., Zweidler-McKay,P.A., Tong,B., and Tsichlis,P.N. 1996a. The Gfi-1 proto-oncoprotein contains a novel transcriptional repressor domain, SNAG, and inhibits G1 arrest induced by interleukin-2 withdrawal. Mol Cell Biol. 16: 6263-6272.

Grimes,H.L., Gilks,C.B., Chan,T.O., Porter,S., and Tsichlis,P.N. 1996b. The Gfi-1 protooncoprotein represses Bax expression and inhibits T-cell death. Proc. Natl. Acad. Sci. U. S. A 93: 14569-14573.

Grinberg,I. and Millen,K.J. 2005. The ZIC gene family in development and disease. Clin. Genet. 67: 290-296.

Grompe,M. 1993. The rapid detection of unknown mutations in nucleic acids. Nat. Genet. 5: 111-117.

Gross,E., Arnold,N., Goette,J., Schwarz-Boeger,U., and Kiechle,M. 1999. A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC. Hum. Genet. 105: 72-78.

Gruber,C.A., Rhee,J.M., Gleiberman,A., and Turner,E.E. 1997. POU domain factors of the Brn-3 class recognize functional DNA elements which are distinctive, symmetrical, and highly conserved in evolution. Mol Cell Biol. 17: 2391-2400.

Gut,I.G. 2001. Automation in genotyping of single nucleotide polymorphisms. Hum. Mutat. 17: 475-492.

Gutierrez,B., Arias,B., Gasto,C., Catalan,R., Papiol,S., Pintor,L., and Fananas,L. 2004. Association analysis between a functional polymorphism in the monoamine oxidase A gene promoter and severe mood disorders. Psychiatr. Genet. 14: 203-208.

Hagen,G., Muller,S., Beato,M., and Suske,G. 1992. Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. Nucleic Acids Res. 20: 5519-5525.

Hagen,G., Muller,S., Beato,M., and Suske,G. 1994. Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J. 13: 3843-3851.

Haider,N.B., Ikeda,A., Naggert,J.K., and Nishina,P.M. 2002. Genetic modifiers of vision and hearing. Hum. Mol Genet. 11: 1195-1206.

Hall,J.M., Lee,M.K., Newman,B., Morrow,J.E., Anderson,L.A., Huey,B., and King,M.C. 1990. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250: 1684-1689.

Hanahan.D., 1985. Cloning: a practical approach. Techniques for transformation in E.Coli. IRL Press, Oxford, Washington.

Hata,Y., Duh,E., Zhang,K., Robinson,G.S., and Aiello,L.P. 1998. Transcription factors Sp1 and Sp3 alter vascular endothelial growth factor receptor expression through a novel recognition sequence. J. Biol. Chem. 273: 19294-19303.

Hauge,X.Y. and Litt,M. 1993. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. Hum. Mol Genet. 2: 411-415.

Hawkins,R.D. and Lovett,M. 2004. The developmental genetics of auditory hair cells. Hum. Mol Genet. 13 Spec No 2: R289-R296.

Hayashi,K. and Yandell,D.W. 1993. How sensitive is PCR-SSCP? Hum. Mutat. 2: 338-346.

He,X., Treacy,M.N., Simmons,D.M., Ingraham,H.A., Swanson,L.W., and Rosenfeld,M.G. 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. Nature 340: 35-41.

Heller,S., Sheane,C.A., Javed,Z., and Hudspeth,A.J. 1998. Molecular markers for cell types of the inner ear and candidate genes for hearing disorders. Proc. Natl. Acad. Sci. U. S. A 95: 11400-11405.

Hennessy,L.K., Teare,J., and Ko,C. 1998. PCR conditions and DNA denaturants affect reproducibility of single-strand conformation polymorphism patterns for BRCA1 mutations. Clin. Chem. 44: 879-882.

Her,S., Claycomb,R., Tai,T.C., and Wong,D.L. 2003. Regulation of the rat phenylethanolamine N-methyltransferase gene by transcription factors Sp1 and MAZ. Mol Pharmacol 64: 1180-1188.

Herrmann,S.M. and Paul,M. 2002. Studying genotype-phenotype relationships: cardiovascular disease as an example. J. Mol Med. 80: 282-289.

Hertzano,R., Montcouquiol,M., Rashi-Elkeles,S., Elkon,R., Yucel,R., Frankel,W.N., Rechavi,G., Moroy,T., Friedman,T.B., Kelley,M.W., and Avraham,K.B. 2004. Transcription profiling of inner ears from Pou4f3(ddl/ddl) identifies Gfi1 as a target of the Pou4f3 deafness gene. Hum. Mol Genet. 13: 2143-2153.

Hirano,F., Tanaka,H., Hirano,Y., Hiramoto,M., Handa,H., Makino,I., and Scheidereit,C. 1998. Functional interference of Sp1 and NF-kappaB through the same DNA binding site. Mol Cell Biol. 18: 1266-1274.

Hirschhorn,J.N. and Altshuler,D. 2002. Once and again-issues surrounding replication in genetic association studies. J. Clin. Endocrinol. Metab 87: 4438-4441.

Hirschhorn,J.N., Lohmueller,K., Byrne,E., and Hirschhorn,K. 2002. A comprehensive review of genetic association studies. Genet. Med. 4: 45-61.

Hirschhorn,J.N. and Daly,M.J. 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6: 95-108.

Hock,H., Hamblen,M.J., Rooke,H.M., Traver,D., Bronson,R.T., Cameron,S., and Orkin,S.H. 2003. Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. Immunity. 18: 109-120.

Holley,M.C. 2005. Keynote review: The auditory system, hearing loss and potential targets for drug development. Drug Discov. Today 10: 1269-1282.

Holme,R.H. and Steel,K.P. 2004. Progressive hearing loss and increased susceptibility to noise-induced hearing loss in mice carrying a Cdh23 but not a Myo7a mutation. J. Assoc. Res. Otolaryngol. 5: 66-79.

Holt,J.R. and Corey,D.P. 2000. Two mechanisms for transducer adaptation in vertebrate hair cells. Proc. Natl. Acad. Sci. U. S. A 97: 11730-11735.

Hong,O.S. and Kim,M.J. 2001. Factors associated with hearing loss among workers of the airline industry in Korea. ORL Head Neck Nurs. 19: 7-13.

Hoogendoorn,B., Coleman,S.L., Guy,C.A., Smith,K., Bowen,T., Buckland,P.R., and O'Donovan,M.C. 2003. Functional analysis of human promoter polymorphisms. Hum. Mol Genet. 12: 2249-2254.

Hoppe,K.L. and Francone,O.L. 1998. Binding and functional effects of transcription factors Sp1 and Sp3 on the proximal human lecithin:cholesterol acyltransferase promoter. J. Lipid Res. 39: 969-977.

Howard,J. and Hudspeth,A.J. 1988. Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the bullfrog's saccular hair cell. Neuron 1: 189-199.

Hu,Z., Miao,X., Ma,H., Wang,X., Tan,W., Wei,Q., Lin,D., and Shen,H. 2005. A common polymorphism in the 3'UTR of cyclooxygenase 2/prostaglandin synthase 2 gene and risk of lung cancer in a Chinese population. Lung Cancer 48: 11-17.

Hudspeth,A.J. 1997. How hearing happens. Neuron 19: 947-950.

Hudspeth,A.J. 2005. How the ear's works work: mechanoelectrical transduction and amplification by hair cells. C. R. Biol. 328: 155-162.

Hultcrantz,M., Sylven,L., and Borg,E. 1994. Ear and hearing problems in 44 middle-aged women with Turner's syndrome. Hear. Res. 76: 127-132.

Hultcrantz,M., Stenberg,A.E., Fransson,A., and Canlon,B. 2000. Characterization of hearing in an X,0 'Turner mouse'. Hear. Res. 143: 182-188.

Humphries,S.E., Ridker,P.M., and Talmud,P.J. 2004. Genetic testing for cardiovascular disease susceptibility: a useful clinical management tool or possible misinformation? Arterioscler. Thromb. Vasc. Biol. 24: 628-636.

Iglesias,A.R., Kindlund,E., Tammi,M., and Wadelius,C. 2004. Some microsatellites may act as novel polymorphic cis-regulatory elements through transcription factor binding. Gene 341: 149-165.

International HapMap Consortium, 2003. The International HapMap Project. Nature. 426: 789-796.

Ioannidis,J.P., Trikalinos,T.A., Ntzani,E.E., and Contopoulos-Ioannidis,D.G. 2003. Genetic associations in large versus small studies: an empirical assessment. Lancet 361: 567-571.

Ishii,N., Wanaka,A., Ohno,K., Matsumoto,K., Eguchi,Y., Mori,T., Tsujimoto,Y., and Tohyama,M. 1996. Localization of bcl-2, bax, and bcl-x mRNAs in the developing inner ear of the mouse. Brain Res. 726: 123-128.

Iype,T., Francis,J., Garmey,J.C., Schisler,J.C., Nesher,R., Weir,G.C., Becker,T.C., Newgard,C.B., Griffen,S.C., and Mirmira,R.G. 2005. Mechanism of insulin gene regulation by the pancreatic transcription factor Pdx-1: application of pre-mRNA analysis and chromatin immunoprecipitation to assess formation of functional transcriptional complexes. J. Biol. Chem. 280: 16798-16807.

Izumikawa,M., Minoda,R., Kawamoto,K., Abrashkin,K.A., Swiderski,D.L., Dolan,D.F., Brough,D.E., and Raphael,Y. 2005. Auditory hair cell replacement and hearing improvement by Atoh1 gene therapy in deaf mammals. Nat. Med. 11: 271-276.

Jafar-Nejad,H. and Bellen,H.J. 2004. Gfi/Pag-3/senseless zinc finger proteins: a unifying theme? Mol Cell Biol. 24: 8803-8812.

Jain,P.K., Lalwani,A.K., Li,X.C., Singleton,T.L., Smith,T.N., Chen,A., Deshmukh,D., Verma,I.C., Smith,R.J., and Wilcox,E.R. 1998. A gene for recessive nonsyndromic sensorineural deafness (DFNB18) maps to the chromosomal region 11p14-p15.1 containing the Usher syndrome type 1C gene. Genomics 50: 290-292.

Jang,J.H. and Surh,Y.J. 2003. Potentiation of cellular antioxidant capacity by Bcl-2: implications for its antiapoptotic function. Biochem. Pharmacol 66: 1371-1379.

Jat,P.S., Noble,M.D., Ataliotis,P., Tanaka,Y., Yannoutsos,N., Larsen,L., and Kioussis,D. 1991. Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse. Proc. Natl. Acad. Sci. U. S. A 88: 5096-5100.

Jennings,C.R. and Jones,N.S. 2001. Presbyacusis. J. Laryngol. Otol. 115: 171-178.

Jensen,O.N., Podtelejnikov,A., and Mann,M. 1996. Delayed extraction improves specificity in database searches by matrix-assisted laser desorption/ionization peptide maps. Rapid Commun. Mass Spectrom. 10: 1371-1378.

Johnson,G.C., Esposito,L., Barratt,B.J., Smith,A.N., Heward,J., Di Genova,G., Ueda,H., Cordell,H.J., Eaves,I.A., Dudbridge,F., Twells,R.C., Payne,F., Hughes,W., Nutland,S., Stevens,H., Carr,P., Tuomilehto-Wolf,E., Tuomilehto,J., Gough,S.C., Clayton,D.G., and Todd,J.A. 2001. Haplotype tagging for the identification of common disease genes. Nat. Genet. 29: 233-237.

Johnson,K.R., Erway,L.C., Cook,S.A., Willott,J.F., and Zheng,Q.Y. 1997. A major gene affecting age-related hearing loss in C57BL/6J mice. Hear. Res. 114: 83-92.

Johnson,K.R., Zheng,Q.Y., and Erway,L.C. 2000. A major gene affecting age-related hearing loss is common to at least ten inbred strains of mice. Genomics 70: 171-180.

Johnson,K.R. and Zheng,Q.Y. 2002. Ahl2, a second locus affecting age-related hearing loss in mice. Genomics 80: 461-464.

Johnson,K.R., Zheng,Q.Y., and Noben-Trauth,K. 2006. Strain background effects and genetic modifiers of hearing in mice. Brain Res. 1091: 79-88.

Jonsson,R., Rosenhall,U., Gause-Nilsson,I., and Steen,B. 1998. Auditory function in 70- and 75-year-olds of four age cohorts. A cross-sectional and time-lag study of presbyacusis. Scand. Audiol. 27: 81-93.

Jormsjo,S., Ye,S., Moritz,J., Walter,D.H., Dimmeler,S., Zeiher,A.M., Henney,A., Hamsten,A., and Eriksson,P. 2000. Allele-specific regulation of matrix metalloproteinase-12 gene activity is associated with coronary artery luminal dimensions in diabetic patients with manifest coronary artery disease. Circ. Res. 86: 998-1003.

Jugessur,A., Frost,P., Andersen,T.I., Steine,S., Lindblom,A., Borresen-Dale,A.L., and Eiken,H.G. 2000. Enhanced detection of mutations in BRCA1 exon 11 using restriction endonuclease fingerprinting-single-strand conformation polymorphism. J. Mol Med. 78: 580-587.

Kachar,B., Parakkal,M., Kurc,M., Zhao,Y., and Gillespie,P.G. 2000. High-resolution structure of hair-cell tip links. Proc. Natl. Acad. Sci. U. S. A 97: 13336-13341.

Kaczynski,J., Cook,T., and Urrutia,R. 2003. Sp1- and Kruppel-like transcription factors. Genome Biol. 4: 206.

Kalatzis,V. and Petit,C. 1998. The fundamental and medical impacts of recent progress in research on hereditary hearing loss. Hum. Mol Genet. 7: 1589-1597.

Karlsson,K.K., Harris,J.R., and Svartengren,M. 1997. Description and primary results from an audiometric study of male twins. Ear Hear. 18: 114-120.

Karsunky,H., Zeng,H., Schmidt,T., Zevnik,B., Kluge,R., Schmid,K.W., Duhrsen,U., and Moroy,T. 2002. Inflammatory reactions and severe neutropenia in mice lacking the transcriptional repressor Gfi1. Nat. Genet. 30: 295-300.

Kashi,Y., King,D., and Soller,M. 1997. Simple sequence repeats as a source of quantitative genetic variation. Trends Genet. 13: 74-78.

Keats,B.J. and Berlin,C.I. 1999. Genomics and hearing impairment. Genome Res. 9: 7-16.

Keavney,B., McKenzie,C., Parish,S., Palmer,A., Clark,S., Youngman,L., Delepine,M., Lathrop,M., Peto,R., and Collins,R. 2000. Large-scale test of hypothesised associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. Lancet 355: 434-442.

Keithley,E.M., Erkman,L., Bennett,T., Lou,L., and Ryan,A.F. 1999. Effects of a hair cell transcription factor, Brn-3.1, gene deletion on homozygous and heterozygous mouse cochleas in adulthood and aging. Hear. Res. 134: 71-76.

Kiernan,A.E. and Steel,K.P. 2000. Mouse homologues for human deafness. Adv. Otorhinolaryngol. 56: 233-243.

Kingsley,C. and Winoto,A. 1992. Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. Mol Cell Biol. 12: 4251-4261.

Kingston,R.E., Bunker,C.A., and Imbalzano,A.N. 1996. Repression and activation by multiprotein complexes that alter chromatin structure. Genes Dev. 10: 905-920.

Knapp,L.A. 2005. Denaturing gradient gel electrophoresis and its use in the detection of major histocompatibility complex polymorphism. Tissue Antigens 65: 211-219.

Knight,J.C. 2003. Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. Clin. Sci. (Lond) 104: 493-501.

Kodama,D., Saito,M., Matsumoto,W., Sabouri,A.H., Izumo,S., Arimura,K., Usuku,K., Bangham,C.R., and Osame,M. 2004. Longer dinucleotide repeat polymorphism in matrix metalloproteinase-9 (MMP-9) gene promoter which correlates with higher HTLV-I Tax mediated transcriptional activity influences the risk of HTLV-I associated myelopathy/tropical spastic paraparesis (HAM/TSP). J. Neuroimmunol. 156: 188-194.

Kopke,R., Allen,K.A., Henderson,D., Hoffer,M., Frenz,D., and Van de,W.T. 1999. A radical demise. Toxins and trauma share common pathways in hair cell death. Ann. N. Y. Acad. Sci. 884: 171-191.

Koritschoner,N.P., Bocco,J.L., Panzetta-Dutari,G.M., Dumur,C.I., Flury,A., and Patrito,L.C. 1997. A novel human zinc finger protein that interacts with the core promoter element of a TATA box-less gene. J. Biol. Chem. 272: 9573-9580.

Koutsodontis,G., Tentes,I., Papakosta,P., Moustakas,A., and Kardassis,D. 2001. Sp1 plays a critical role in the transcriptional activation of the human cyclin-dependent kinase inhibitor p21(WAF1/Cip1) gene by the p53 tumor suppressor protein. J. Biol. Chem. 276: 29116-29125.

Kozel,P.J., Davis,R.R., Krieg,E.F., Shull,G.E., and Erway,L.C. 2002. Deficiency in plasma membrane calcium ATPase isoform 2 increases susceptibility to noise-induced hearing loss in mice. Hear. Res. 164: 231-239.

Kubisch,C., Schroeder,B.C., Friedrich,T., Lutjohann,B., El Amraoui,A., Marlin,S., Petit,C., and Jentsch,T.J. 1999. KCNQ4, a novel potassium channel expressed in sensory outer hair cells, is mutated in dominant deafness. Cell 96: 437-446.

Kujoth,G.C., Hiona,A., Pugh,T.D., Someya,S., Panzer,K., Wohlgemuth,S.E., Hofer,T., Seo,A.Y., Sullivan,R., Jobling,W.A., Morrow,J.D., Van Remmen,H., Sedivy,J.M., Yamasoba,T., Tanokura,M., Weindruch,R., Leeuwenburgh,C., and Prolla,T.A. 2005. Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. Science 309: 481-484.

Kukita,Y., Tahira,T., Sommer,S.S., and Hayashi,K. 1997. SSCP analysis of long DNA fragments in low pH gel. Hum. Mutat. 10: 400-407.

Kunkel,T.A. and Bebenek,K. 2000. DNA replication fidelity. Annu. Rev. Biochem. 69: 497-529.

Kunkel,T.A. 2004. DNA replication fidelity. J. Biol. Chem. 279: 16895-16898.

Kwan,K.Y., Allchorne,A.J., Vollrath,M.A., Christensen,A.P., Zhang,D.S., Woolf,C.J., and Corey,D.P. 2006. TRPA1 contributes to cold, mechanical, and chemical nociception but is not essential for hair-cell transduction. Neuron 50: 277-289.

Lagziel,A., Ahmed,Z.M., Schultz,J.M., Morell,R.J., Belyantseva,I.A., and Friedman,T.B. 2005. Spatiotemporal pattern and isoforms of cadherin 23 in wild type and waltzer mice during inner ear hair cell development. Dev. Biol. 280: 295-306.

Lambert,J.C., Luedecking-Zimmer,E., Merrot,S., Hayes,A., Thaker,U., Desai,P., Houzet,A., Hermant,X., Cottel,D., Pritchard,A., Iwatsubo,T., Pasquier,F., Frigard,B., Conneally,P.M., Chartier-Harlin,M.C., DeKosky,S.T., Lendon,C., Mann,D., Kamboh,M.I., and Amouyel,P. 2003. Association of 3'-UTR polymorphisms of the oxidised LDL receptor 1 (OLR1) gene with Alzheimer's disease. J. Med. Genet. 40: 424-430.

Landen,D., Wilkins,S., Stephenson,M., and McWilliams,L. 2004. Noise exposure and hearing loss among sand and gravel miners. J. Occup. Environ. Hyg. 1: 532-541.

Lander,E.S. and Schork,N.J. 1994. Genetic dissection of complex traits. Science 265: 2037-2048.

Lanford,P.J., Lan,Y., Jiang,R., Lindsell,C., Weinmaster,G., Gridley,T., and Kelley,M.W. 1999. Notch signalling pathway mediates hair cell development in mammalian cochlea. Nat. Genet. 21: 289-292.

Latchman,D.S. 1998. The Brn-3a transcription factor. Int. J. Biochem. Cell Biol. 30: 1153-1157.

Latchman,D.S. 1999. POU family transcription factors in the nervous system. J. Cell Physiol 179: 126-133.

Lawrence,R.W., Evans,D.M., and Cardon,L.R. 2005. Prospects and pitfalls in whole genome association studies. Philos. Trans. R. Soc. Lond B Biol. Sci. 360: 1589-1595.

Le Prell,C.G., Dolan,D.F., Schacht,J., Miller,J.M., Lomax,M.I., and Altschuler,R.A. 2003. Pathways for protection from noise induced hearing loss. Noise. Health 5: 1-17.

Lee,K.J. 1999. Essential Otolaryngology Head and Neck Surgery. Stamford Connecticut: Appleton and Lange.

Lee,T.I. and Young,R.A. 2000. Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. 34: 77-137.

Lee,Y.H., Yano,M., Liu,S.Y., Matsunaga,E., Johnson,P.F., and Gonzalez,F.J. 1994. A novel cis-acting element controlling the rat CYP2D5 gene and requiring cooperativity between C/EBP beta and an Sp1 factor. Mol Cell Biol. 14: 1383-1394.

Lefebvre,P.P., Malgrange,B., Lallemend,F., Staecker,H., Moonen,G., and Van De Water,T.R. 2002. Mechanisms of cell death in the injured auditory system: otoprotective strategies. Audiol. Neurootol. 7: 165-170.

LeMasurier,M. and Gillespie,P.G. 2005. Hair-cell mechanotransduction and cochlear amplification. Neuron 48: 403-415.

Levinson,G. and Gutman,G.A. 1987a. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol. Evol. 4: 203-221.

Levinson,G. and Gutman,G.A. 1987b. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. Nucleic Acids Res. 15: 5323-5338.

Li,S., Price,S.M., Cahill,H., Ryugo,D.K., Shen,M.M., and Xiang,M. 2002. Hearing loss caused by progressive degeneration of cochlear hair cells in mice deficient for the Barhl1 homeobox gene. Development 129: 3523-3532.

Li,Y.C., Korol,A.B., Fahima,T., Beiles,A., and Nevo,E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol. 11: 2453-2465.

Li,Y.C., Korol,A.B., Fahima,T., and Nevo,E. 2004. Microsatellites within genes: structure, function, and evolution. Mol Biol. Evol. 21: 991-1007.

Liberman,M.C., Gao,J., He,D.Z., Wu,X., Jia,S., and Zuo,J. 2002. Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. Nature 419: 300-304.

Lietz,M., Hohl,M., and Thiel,G. 2003. RE-1 silencing transcription factor (REST) regulates human synaptophysin gene transcription through an intronic sequence-specific DNA-binding site. Eur. J. Biochem. 270: 2-9.

Lillycrop,K.A., Budrahan,V.S., Lakin,N.D., Terrenghi,G., Wood,J.N., Polak,J.M., and Latchman,D.S. 1992. A novel POU family transcription factor is closely related to Brn-3 but has a distinct expression pattern in neuronal cells. Nucleic Acids Res. 20: 5093-5096.

Lin,S.Y., Black,A.R., Kostic,D., Pajovic,S., Hoover,C.N., and Azizkhan,J.C. 1996. Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction. Mol Cell Biol. 16: 1668-1675.

Litt,M., Hauge,X., and Sharma,V. 1993. Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. Biotechniques 15: 280-284.

Liu,Q. and Sommer,S.S. 1995. Restriction endonuclease fingerprinting (REF): a sensitive method for screening mutations in long, contiguous segments of DNA. Biotechniques 18: 470-477.

Liu,Y.Z., Dawson,S.J., and Latchman,D.S. 1996. Alternative splicing of the Brn-3a and Brn-3b transcription factor RNAs is regulated in neuronal cells. J. Mol Neurosci. 7: 77-85.

Lohmueller,K.E., Pearce,C.L., Pike,M., Lander,E.S., and Hirschhorn,J.N. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat. Genet. 33: 177-182.

Lopez-Rodriguez,C., Botella,L., and Corbi,A.L. 1997. CCAAT-enhancer-binding proteins (C/EBP) regulate the tissue specific activity of the CD11c integrin gene promoter through functional interactions with Sp1 proteins. J. Biol. Chem. 272: 29120-29126.

Lynch,E.D. and Kil,J. 2005. Compounds for the prevention and treatment of noise-induced hearing loss. Drug Discov. Today 10: 1291-1298.

Majewski,J. and Ott,J. 2000. GT repeats are associated with recombination on human chromosome 22. Genome Res. 10: 1108-1114.

Majores,M., Kolsch,H., Bagli,M., Papassotiropoulos,A., Lohmann,P.L., Schmitz,S., Rao,M.L., Maier,W., and Heun,R. 2002. Cathepsin D: screening for new polymorphisms using single-strand conformation polymorphism analysis. Int. J. Mol Med. 9: 185-187.

Malgrange,B., Rigo,J.M., Coucke,P., Thiry,M., Hans,G., Nguyen,L., Van De Water,T.R., Moonen,G., and Lefebvre,P.P. 2002. Identification of factors that maintain mammalian outer hair cells in adult organ of Corti explants. Hear. Res. 170: 48-58.

Malik,K.F., Kim,J., Hartman,A.L., Kim,P., and Young,W.S., III 1996. Binding preferences of the POU domain protein Brain-4: implications for autoregulation. Brain Res. Mol Brain Res. 38: 209-221.

Marzella,P.L. and Gillespie,L.N. 2002. Role of trophic factors in the development, survival and repair of primary auditory neurons. Clin. Exp. Pharmacol Physiol 29: 363-371.

Mason,J.A. and Herrmann,K.R. 1998. Universal infant hearing screening by automated auditory brainstem response measurement. Pediatrics 101: 221-228.

Mastrangelo,I.A., Courey,A.J., Wall,J.S., Jackson,S.P., and Hough,P.V. 1991. DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. Proc. Natl. Acad. Sci. U. S. A 88: 5670-5674.

Matsui,J.I., Parker,M.A., Ryals,B.M., and Cotanche,D.A. 2005. Regeneration and replacement in the vertebrate inner ear. Drug Discov. Today 10: 1307-1312.

Mburu,P., Mustapha,M., Varela,A., Weil,D., El Amraoui,A., Holme,R.H., Rump,A., Hardisty,R.E., Blanchard,S., Coimbra,R.S., Perfettini,I., Parkinson,N., Mallon,A.M., Glenister,P., Rogers,M.J., Paige,A.J., Moir,L., Clay,J., Rosenthal,A., Liu,X.Z., Blanco,G., Steel,K.P., Petit,C., and Brown,S.D. 2003. Defects in whirlin, a PDZ domain molecule involved in stereocilia elongation, cause deafness in the whirler mouse and families with DFNB31. Nat. Genet. 34: 421-428.

McCord,J.M. and Fridovich,I. 1969. Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). J. Biol. Chem. 244: 6049-6055.

McEvilly,R.J., Erkman,L., Luo,L., Sawchenko,P.E., Ryan,A.F., and Rosenfeld,M.G. 1996. Requirement for Brn-3.0 in differentiation and survival of sensory and motor neurons. Nature 384: 574-577.

McFadden,S.L., Ding,D., Burkard,R.F., Jiang,H., Reaume,A.G., Flood,D.G., and Salvi,R.J. 1999a. Cu/Zn SOD deficiency potentiates hearing loss and cochlear pathology in aged 129,CD-1 mice. J. Comp Neurol. 413: 101-112.

McFadden,S.L., Ding,D., Reaume,A.G., Flood,D.G., and Salvi,R.J. 1999b. Age-related cochlear hair cell loss is enhanced in mice lacking copper/zinc superoxide dismutase. Neurobiol. Aging 20: 1-8.

McFadden,S.L., Ohlemiller,K.K., Ding,D., Shero,M., and Salvi,R.J. 2001. The Influence of Superoxide Dismutase and Glutathione Peroxidase Deficiencies on Noise-Induced Hearing Loss in Mice. Noise. Health 3: 49-64.

McGhee,L., Bryan,J., Elliott,L., Grimes,H.L., Kazanjian,A., Davis,J.N., and Meyers,S. 2003. Gfi-1 attaches to the nuclear matrix, associates with ETO (MTG8) and histone deacetylase proteins, and represses transcription using a TSA-sensitive mechanism. J. Cell Biochem. 89: 1005-1018.

Menger,D.J. and Tange,R.A. 2003. The aetiology of otosclerosis: a review of the literature. Clin. Otolaryngol. Allied Sci. 28: 112-120.

Michaud,J., Brody,L.C., Steel,G., Fontaine,G., Martin,L.S., Valle,D., and Mitchell,G. 1992. Strand-separating conformational polymorphism analysis: efficacy of detection of point mutations in the human ornithine delta-aminotransferase gene. Genomics 13: 389-394.

Michel,V., Goodyear,R.J., Weil,D., Marcotti,W., Perfettini,I., Wolfrum,U., Kros,C.J., Richardson,G.P., and Petit,C. 2005. Cadherin 23 is a component of the transient lateral links in the developing hair bundles of cochlear sensory cells. Dev. Biol. 280: 281-294.

Miki,Y., Swensen,J., Shattuck-Eidens,D., Futreal,P.A., Harshman,K., Tavtigian,S., Liu,Q., Cochran,C., Bennett,L.M., Ding,W., and . 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266: 66-71.

Miller,J.M., Dolan,D.F., Raphael,Y., and Altschuler,R.A. 1998. Interactive effects of aging with noise induced hearing loss. Scand. Audiol. Suppl 48: 53-61.

Misquitta,C.M., Iyer,V.R., Werstiuk,E.S., and Grover,A.K. 2001. The role of 3'-untranslated region (3'-UTR) mediated mRNA stability in cardiovascular pathophysiology. Mol Cell Biochem. 224: 53-67.

Moore,K.L. and Dalley,A.F. 1999. Clinically Orientated Anatomy. Baltimore: Lippincott, Williams and Wilkins.

Moroy,T. 2005. The zinc finger transcription factor Growth factor independence 1 (Gfi1). Int. J. Biochem. Cell Biol. 37: 541-546.

Morrison,A., Hodgetts,C., Gossler,A., Hrabe,d.A., and Lewis,J. 1999. Expression of Delta1 and Serrate1 (Jagged1) in the mouse inner ear. Mech. Dev. 84: 169-172.

Morton,N.E. 1991. Genetic epidemiology of hearing impairment. Ann. N. Y. Acad. Sci. 630: 16-31.

Moskvina,V., Holmans,P., Schmidt,K.M., and Craddock,N. 2005. Design of case-controls studies with unscreened controls. Ann. Hum. Genet. 69: 566-576.

Nataraj,A.J., Olivos-Glander,I., Kusukawa,N., and Highsmith,W.E., Jr. 1999. Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis 20: 1177-1185.

Naylor,L.H. and Clark,E.M. 1990. d(TG)n.d(CA)n sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. Nucleic Acids Res. 18: 1595-1601.

Nelson,D.I., Nelson,R.Y., Concha-Barrientos,M., and Fingerhut,M. 2005. The global burden of occupational noise-induced hearing loss. Am. J. Ind. Med. 48: 446-458.

Nemoto,M., Morita,Y., Mishima,Y., Takahashi,S., Nomura,T., Ushiki,T., Shiroishi,T., Kikkawa,Y., Yonekawa,H., and Kominami,R. 2004. Ahl3, a third locus on mouse chromosome 17 affecting age-related hearing loss. Biochem. Biophys. Res. Commun. 324: 1283-1288.

Niimi,T., Munakata,M., Keck-Waggoner,C.L., Popescu,N.C., Levitt,R.C., Hisada,M., and Kimura,S. 2002. A polymorphism in the human UGRP1 gene promoter that regulates transcription is associated with an increased risk of asthma. Am. J. Hum. Genet. 70: 718-725.

Nissim-Rafinia,M. and Kerem,B. 2002. Splicing regulation as a potential genetic modifier. Trends Genet. 18: 123-127.

Noben-Trauth,K., Zheng,Q.Y., Johnson,K.R., and Nishina,P.M. 1997. mdfw: a deafness susceptibility locus that interacts with deaf waddler (dfw). Genomics 44: 266-272.

Noben-Trauth,K., Zheng,Q.Y., and Johnson,K.R. 2003. Association of cadherin 23 with polygenic inheritance and genetic modification of sensorineural hearing loss. Nat. Genet. 35: 21-23.

Nobili,R., Mammano,F., and Ashmore,J. 1998. How well do we understand the cochlea? Trends Neurosci. 21: 159-167.

Nolan,P.M., Peters,J., Strivens,M., Rogers,D., Hagan,J., Spurr,N., Gray,I.C., Vizor,L., Brooker,D., Whitehill,E., Washbourne,R., Hough,T., Greenaway,S., Hewitt,M., Liu,X., McCormack,S., Pickford,K., Selley,R., Wells,C., Tymowska-Lalanne,Z., Roby,P., Glenister,P., Thornton,C., Thaung,C., Stevenson,J.A., Arkell,R., Mburu,P., Hardisty,R., Kiernan,A., Erven,A., Steel,K.P., Voegeling,S., Guenet,J.L., Nickols,C., Sadri,R., Nasse,M., Isaacs,A., Davies,K., Browne,M., Fisher,E.M., Martin,J., Rastan,S., Brown,S.D., and Hunter,J. 2000. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat. Genet. 25: 440-443.

Nordt,T.K., Lohrmann,J., and Bode,C. 2001. Regulation of PAI-1 expression by genetic polymorphisms. Impact on atherogenesis. Thromb. Res. 103 Suppl 1: S1-S5.

O'Donovan,K.J., Tourtellotte,W.G., Millbrandt,J., and Baraban,J.M. 1999. The EGR family of transcription-regulatory factors: progress at the interface of molecular and systems neuroscience. Trends Neurosci. 22: 167-173.

Ohinata,Y., Miller,J.M., and Schacht,J. 2003. Protection from noise-induced lipid peroxidation and hair cell loss in the cochlea. Brain Res. 966: 265-273.

Ohlemiller,K.K., McFadden,S.L., Ding,D.L., Flood,D.G., Reaume,A.G., Hoffman,E.K., Scott,R.W., Wright,J.S., Putcha,G.V., and Salvi,R.J. 1999. Targeted deletion of the cytosolic Cu/Zn-superoxide dismutase gene (Sod1) increases susceptibility to noise-induced hearing loss. Audiol. Neurootol. 4: 237-246.

Ohlemiller,K.K., McFadden,S.L., Ding,D.L., Lear,P.M., and Ho,Y.S. 2000. Targeted mutation of the gene for cellular glutathione peroxidase (Gpx1) increases noise-induced hearing loss in mice. J. Assoc. Res. Otolaryngol. 1: 243-254.

Ohlemiller,K.K. 2004. Age-related hearing loss: the status of Schuknecht's typology. Curr. Opin. Otolaryngol. Head Neck Surg. 12: 439-443.

Ohlemiller,K.K. and Gagnon,P.M. 2004. Cellular correlates of progressive hearing loss in 129S6/SvEv mice. J. Comp Neurol. 469: 377-390.

Ohtsuka,T., Ishibashi,M., Gradwohl,G., Nakanishi,S., Guillemot,F., and Kageyama,R. 1999. Hes1 and Hes5 as notch effectors in mammalian neuronal differentiation. EMBO J. 18: 2196-2207.

Orita,M., Iwahana,H., Kanazawa,H., Hayashi,K., and Sekiya,T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. U. S. A 86: 2766-2770.

Osawa,M., Yamaguchi,T., Nakamura,Y., Kaneko,S., Onodera,M., Sawada,K., Jegalian,A., Wu,H., Nakauchi,H., and Iwama,A. 2002. Erythroid expansion mediated by the Gfi-1B zinc finger protein: role in normal hematopoiesis. Blood 100: 2769-2777.

Ozaki,K., Ohnishi,Y., Iida,A., Sekine,A., Yamada,R., Tsunoda,T., Sato,H., Sato,H., Hori,M., Nakamura,Y., and Tanaka,T. 2002. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat. Genet. 32: 650-654.

Page,G.P. and Amos,C.I. 1999. Comparison of linkage-disequilibrium methods for localization of genes influencing quantitative traits in humans. Am. J. Hum. Genet. 64: 1194-1205.

Pan,L., Yang,Z., Feng,L., and Gan,L. 2005. Functional equivalence of Brn3 POU-domain transcription factors in mouse retinal neurogenesis. Development 132: 703-712.

Parham,K. 1997. Distortion product otoacoustic emissions in the C57BL/6J mouse model of age-related hearing loss. Hear. Res. 112: 216-234.

Parkinson,N. and Brown,S.D. 2002. Focusing on the genetics of hearing: you ain't heard nothin' yet. Genome Biol. 3: COMMENT2006.

Partridge,L. and Gems,D. 2002. Mechanisms of ageing: public or private? Nat. Rev. Genet. 3: 165-175.

Parving,A. 1999. The need for universal neonatal hearing screening--some aspects of epidemiology and identification. Acta Paediatr. Suppl 88: 69-72.

Patwardhan,S., Gashler,A., Siegel,M.G., Chang,L.C., Joseph,L.J., Shows,T.B., Le Beau,M.M., and Sukhatme,V.P. 1991. EGR3, a novel member of the Egr family of genes encoding immediate-early transcription factors. Oncogene 6: 917-928.

Perkins,N.D., Agranoff,A.B., Pascal,E., and Nabel,G.J. 1994. An interaction between the DNA-binding domains of RelA(p65) and Sp1 mediates human immunodeficiency virus gene activation. Mol Cell Biol. 14: 6570-6583.

Petersen,M.B. 2002. Non-syndromic autosomal-dominant deafness. Clin. Genet. 62: 1-13.

Petersohn,D. and Thiel,G. 1996. Role of zinc-finger proteins Sp1 and zif268/egr-1 in transcriptional regulation of the human synaptobrevin II gene. Eur. J. Biochem. 239: 827-834.

Petit,C., Levilliers,J., and Hardelin,J.P. 2001. Molecular genetics of hearing loss. Annu. Rev. Genet. 35: 589-646.

Petit,C. 2006. From deafness genes to hearing mechanisms: harmony and counterpoint. Trends Mol Med. 12: 57-64.

Philipsen,S. and Suske,G. 1999. A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. Nucleic Acids Res. 27: 2991-3000.

Phillips,K. and Luisi,B. 2000. The virtuoso of versatility: POU proteins that flex to fit. J. Mol Biol. 302: 1023-1039.

Pickles,J.O., Comis,S.D., and Osborne,M.P. 1984. Cross-links between stereocilia in the guinea pig organ of Corti, and their possible relation to sensory transduction. Hear. Res. 15: 103-112.

Pickles,J.O. 2004. Mutation in mitochondrial DNA as a cause of presbyacusis. Audiol. Neurootol. 9: 23-33.

Platzer,J., Engel,J., Schrott-Fischer,A., Stephan,K., Bova,S., Chen,H., Zheng,H., and Striessnig,J. 2000. Congenital deafness and sinoatrial node dysfunction in mice lacking class D L-type Ca2+ channels. Cell 102: 89-97.

Probst,F.J. and Camper,S.A. 1999. The role of mouse mutants in the identification of human hereditary hearing loss genes. Hear. Res. 130: 1-6.

Quandt,K., Frech,K., Karas,H., Wingender,E., and Werner,T. 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23: 4878-4884.

Rabionet,R., Gasparini,P., and Estivill,X. 2000. Molecular genetics of hearing impairment due to mutations in gap junction genes encoding beta connexins. Hum. Mutat. 16: 190-202.

Ratziu,V., Lalazar,A., Wong,L., Dang,Q., Collins,C., Shaulian,E., Jensen,S., and Friedman,S.L. 1998. Zf9, a Kruppel-like transcription factor up-regulated in vivo during early hepatic fibrosis. Proc. Natl. Acad. Sci. U. S. A 95: 9500-9505.

Ravnik-Glavac,M., Glavac,D., and Dean,M. 1994. Sensitivity of single-strand conformation polymorphism and heteroduplex method for mutation detection in the cystic fibrosis gene. Hum. Mol Genet. 3: 801-807.

Rebbeck,T.R., Spitz,M., and Wu,X. 2004. Assessing the function of genetic variants in candidate gene association studies. Nat. Rev. Genet. 5: 589-597.

Redowicz,M.J. 2002. Myosins and pathology: genetics and biology. Acta Biochim. Pol. 49: 789-804.

Reich,D.E. and Goldstein,D.B. 2001. Detecting association in a case-control study while correcting for population stratification. Genet. Epidemiol. 20: 4-16.

Reich,D.E. and Lander,E.S. 2001. On the allelic spectrum of human disease. Trends Genet. 17: 502-510.

Reiners,J., Nagel-Wolfrum,K., Jurgens,K., Marker,T., and Wolfrum,U. 2006. Molecular basis of human Usher syndrome: deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease. Exp. Eye Res. 83: 97-119.

Resendes,B.L., Williamson,R.E., and Morton,C.C. 2001. At the speed of sound: gene discovery in the auditory system. Am. J. Hum. Genet. 69: 923-935.

Rich,A., Nordheim,A., and Wang,A.H. 1984. The chemistry and biology of left-handed Z-DNA. Annu. Rev. Biochem. 53: 791-846.

Riley,B.B., Chiang,M., Farmer,L., and Heck,R. 1999. The deltaA gene of zebrafish mediates lateral inhibition of hair cells in the inner ear and is regulated by pax2.1. Development 126: 5669-5678.

Risch,N. and Merikangas,K. 1996. The future of genetic studies of complex human diseases. Science 273: 1516-1517.

Risch,N.J. 2000. Searching for genetic determinants in the new millennium. Nature 405: 847-856.

Rivolta,M.N., Grix,N., Lawlor,P., Ashmore,J.F., Jagger,D.J., and Holley,M.C. 1998. Auditory hair cell precursors immortalized from the mammalian inner ear. Proc. Biol. Sci. 265: 1595-1603.

Rivolta,M.N. and Holley,M.C. 2002. Cell lines in inner ear research. J. Neurobiol. 53: 306-318.

Rivolta,M.N., Halsall,A., Johnson,C.M., Tones,M.A., and Holley,M.C. 2002. Transcript profiling of functionally related groups of genes during conditional differentiation of a mammalian cochlear hair cell line. Genome Res. 12: 1091-1099.

Robertson,N.G., Lu,L., Heller,S., Merchant,S.N., Eavey,R.D., McKenna,M., Nadol,J.B., Jr., Miyamoto,R.T., Linthicum,F.H., Jr., Lubianca Neto,J.F., Hudspeth,A.J., Seidman,C.E., Morton,C.C., and Seidman,J.G. 1998. Mutations in a novel cochlear gene cause DFNA9, a human nonsyndromic deafness with vestibular dysfunction. Nat. Genet. 20: 299-303.

Rodel,B., Tavassoli,K., Karsunky,H., Schmidt,T., Bachmann,M., Schaper,F., Heinrich,P., Shuai,K., Elsasser,H.P., and Moroy,T. 2000. The zinc finger protein Gfi-1 can enhance STAT3 signaling by interacting with the STAT3 inhibitor PIAS3. EMBO J. 19: 5845-5855.

Rosenhall,U. 2003. The influence of ageing on noise-induced hearing loss. Noise. Health 5: 47-53.

Rossetti,S., Corra,S., Biasi,M.O., Turco,A.E., and Pignatti,P.F. 1995. Comparison of heteroduplex and single-strand conformation analyses, followed by ethidium fluorescence visualization, for the detection of mutations in four human genes. Mol Cell Probes 9: 195-200.

Rothenburg,S., Koch-Nolte,F., Rich,A., and Haag,F. 2001. A polymorphic dinucleotide repeat in the rat nucleolin gene forms Z-DNA and inhibits promoter activity. Proc. Natl. Acad. Sci. U. S. A 98: 8985-8990.

Ryan,A.K. and Rosenfeld,M.G. 1997. POU domain family values: flexibility, partnerships, and developmental codes. Genes Dev. 11: 1207-1225.

Rybak,L.P. and Whitworth,C.A. 2005. Ototoxicity: therapeutic opportunities. Drug Discov. Today 10: 1313-1321.

Saffer,J.D., Jackson,S.P., and Annarella,M.B. 1991. Developmental expression of Sp1 in the mouse. Mol Cell Biol. 11: 2189-2199.

Sagher,D., Hsu,A., and Strauss,B. 1999. Stabilization of the intermediate in frameshift mutation. Mutat. Res. 423: 73-77.

Saleque,S., Cameron,S., and Orkin,S.H. 2002. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. Genes Dev. 16: 301-306.

Salisbury,B.A., Pungliya,M., Choi,J.Y., Jiang,R., Sun,X.J., and Stephens,J.C. 2003. SNP and haplotype variation in the human genome. Mutat. Res. 526: 53-61.

Sanderson,J.T. and Steel,J. 1967. Noise-induced hearing loss in bench glass-blowers. Ann. Occup. Hyg. 10: 135-141.

Schneider,I. 1972. Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morphol. 27: 353-365.

Scholtz,A.W., Kammen-Jolly,K., Felder,E., Hussl,B., Rask-Andersen,H., and Schrott-Fischer,A. 2001. Selective aspects of human pathology in high-tone hearing loss of the aging inner ear. Hear. Res. 157: 77-86.

Schrijver,I. 2004. Hereditary non-syndromic sensorineural hearing loss: transforming silence to sound. J. Mol Diagn. 6: 275-284.

Schuknecht,H.F. 1955. Presbycusis. Laryngoscope 65: 402-419.

Schuknecht,H.F. 1964. FURTHER OBSERVATIONS ON THE PATHOLOGY OF PRESBYCUSIS. Arch. Otolaryngol. 80: 369-382.

Schuknecht,H.F. and Gacek,M.R. 1993. Cochlear pathology in presbycusis. Ann. Otol. Rhinol. Laryngol. 102: 1-16.

Schultz,J.M., Yang,Y., Caride,A.J., Filoteo,A.G., Penheiter,A.R., Lagziel,A., Morell,R.J., Mohiddin,S.A., Fananapazir,L., Madeo,A.C., Penniston,J.T., and Griffith,A.J. 2005. Modification of human hearing loss by plasma-membrane calcium pump PMCA2. N. Engl. J. Med. 352: 1557-1564.

Scott,D.A., Wang,R., Kreman,T.M., Andrews,M., McDonald,J.M., Bishop,J.R., Smith,R.J., Karniski,L.P., and Sheffield,V.C. 2000. Functional differences of the PDS gene product are associated with phenotypic variation in patients with Pendred syndrome and non-syndromic hearing loss (DFNB4). Hum. Mol Genet. 9: 1709-1715.

Searle,S. and Blackwell,J.M. 1999. Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility. J. Med. Genet. 36: 295-299.

Seidman,M.D., Khan,M.J., Dolan,D.F., and Quirk,W.S. 1996. Age-related differences in cochlear microcirculation and auditory brain stem response. Arch. Otolaryngol. Head Neck Surg. 122: 1221-1226.

Seidman,M.D., Bai,U., Khan,M.J., and Quirk,W.S. 1997. Mitochondrial DNA deletions associated with aging and presbyacusis. Arch. Otolaryngol. Head Neck Surg. 123: 1039-1045.

Seidman,M.D., Khan,M.J., Bai,U., Shirwany,N., and Quirk,W.S. 2000. Biologic activity of mitochondrial metabolites on aging and age-related hearing loss. Am. J. Otol. 21: 161-167.

Seidman,M.D., Ahmad,N., and Bai,U. 2002. Molecular mechanisms of age-related hearing loss. Ageing Res. Rev. 1: 331-343.

Seidman,M.D., Ahmad,N., Joshi,D., Seidman,J., Thawani,S., and Quirk,W.S. 2004. Age-related hearing loss and its association with reactive oxygen species and mitochondrial DNA damage. Acta Otolaryngol. Suppl 16-24.

Seixas,N.S., Goldman,B., Sheppard,L., Neitzel,R., Norton,S., and Kujawa,S.G. 2005. Prospective noise induced changes to hearing among construction industry apprentices. Occup. Environ. Med. 62: 309-317.

Sha,S.H., Taylor,R., Forge,A., and Schacht,J. 2001. Differential vulnerability of basal and apical hair cells is based on intrinsic susceptibility to free radicals. Hear. Res. 155: 1-8.

Shang,Y., Hu,X., DiRenzo,J., Lazar,M.A., and Brown,M. 2000. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell 103: 843-852.

Sharina,I.G., Martin,E., Thomas,A., Uray,K.L., and Murad,F. 2003. CCAAT-binding factor regulates expression of the beta1 subunit of soluble guanylyl cyclase gene in the BE2 human neuroblastoma cell line. Proc. Natl. Acad. Sci. U. S. A 100: 11523-11528.

Sheffield,V.C., Beck,J.S., Kwitek,A.E., Sandstrom,D.W., and Stone,E.M. 1993. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. Genomics 16: 325-332.

Sherf.B.A., Navarro.S.L., Hannah.R.R. and Wood.K.V., 1996. Dual-Luciferase™ Reporter Assay: An advanced co-reporter technology intergrating firefly and *renilla* luciferase assays. Promega Notes Magazine 57: 2-8.

Shifman,S., Bronstein,M., Sternfeld,M., Pisante-Shalom,A., Lev-Lehman,E., Weizman,A., Reznik,I., Spivak,B., Grisaru,N., Karp,L., Schiffer,R., Kotler,M., Strous,R.D., Swartz-Vanetik,M., Knobler,H.Y., Shinar,E., Beckmann,J.S., Yakir,B., Risch,N., Zak,N.B., and Darvasi,A. 2002. A highly significant association between a COMT haplotype and schizophrenia. Am. J. Hum. Genet. 71: 1296-1302.

Shimajiri,S., Arima,N., Tanimoto,A., Murata,Y., Hamada,T., Wang,K.Y., and Sasaguri,Y. 1999. Shortened microsatellite d(CA)21 sequence down-regulates promoter activity of matrix metalloproteinase 9 gene. FEBS Lett. 455: 70-74.

Shinahara,K., Saijo,T., Mori,K., and Kuroda,Y. 2004. Single-strand conformation polymorphism analysis of the FMR1 gene in autistic and mentally retarded children in Japan. J. Med. Invest 51: 52-58.

Shinde,D., Lai,Y., Sun,F., and Arnheim,N. 2003. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Res. 31: 974-980.

Sia,E.A., Kokoska,R.J., Dominska,M., Greenwell,P., and Petes,T.D. 1997. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol Cell Biol. 17: 2851-2858.

Siemens,J., Kazmierczak,P., Reynolds,A., Sticker,M., Littlewood-Evans,A., and Muller,U. 2002. The Usher syndrome proteins cadherin 23 and harmonin form a complex by means of PDZ-domain interactions. Proc. Natl. Acad. Sci. U. S. A 99: 14946-14951.

Siemens,J., Lillo,C., Dumont,R.A., Reynolds,A., Williams,D.S., Gillespie,P.G., and Muller,U. 2004. Cadherin 23 is a component of the tip link in hair-cell stereocilia. Nature 428: 950-955.

Sjottem,E., Andersen,C., and Johansen,T. 1997. Structural and functional analyses of DNA bending induced by Sp1 family transcription factors. J. Mol Biol. 267: 490-504.

Smith,C.L. and Hager,G.L. 1997. Transcriptional regulation of mammalian genes in vivo. A tale of two templates. J. Biol. Chem. 272: 27493-27496.

Smith,D.J. and Lusis,A.J. 2002. The allelic structure of common disease. Hum. Mol Genet. 11: 2455-2461.

Smith,M.D., Dawson,S.J., Boxer,L.M., and Latchman,D.S. 1998a. The N-terminal domain unique to the long form of the Brn-3a transcription factor is essential to protect neuronal cells from apoptosis and for the activation of Bbcl-2 gene expression. Nucleic Acids Res. 26: 4100-4107.

Smith,M.D., Morris,P.J., and Latchman,D.S. 1998b. The Brn-3c transcription factor contains a neuronal-specific activation domain. Neuroreport 9: 851-856.

Smith,M.D., Melton,L.A., Ensor,E.A., Packham,G., Anderson,P., Kinloch,R.A., and Latchman,D.S. 2001. Brn-3a activates the expression of Bcl-x(L) and promotes neuronal survival in vivo as well as in vitro. Mol Cell Neurosci. 17: 460-470.

Sollner,C., Rauch,G.J., Siemens,J., Geisler,R., Schuster,S.C., Muller,U., and Nicolson,T. 2004. Mutations in cadherin 23 affect tip links in zebrafish sensory hair cells. Nature 428: 955-959.

Song,Q., Chao,J., and Chao,L. 1997. DNA polymorphisms in the 5'-flanking region of the human tissue kallikrein gene. Hum. Genet. 99: 727-734.

Spongr,V.P., Flood,D.G., Frisina,R.D., and Salvi,R.J. 1997. Quantitative measures of hair cell loss in CBA and C57BL/6 mice throughout their life spans. J. Acoust. Soc. Am. 101: 3546-3553.

St Jean,P.L., Zhang,X.C., Hart,B.K., Lamlum,H., Webster,M.W., Steed,D.L., Henney,A.M., and Ferrell,R.E. 1995. Characterization of a dinucleotide repeat in the 92 kDa type IV collagenase gene (CLG4B), localization of CLG4B to chromosome 20 and the role of CLG4B in aortic aneurysmal disease. Ann. Hum. Genet. 59 (Pt 1): 17-24.

Staecker,H. and Van De Water,T.R. 1998. Factors controlling hair-cell regeneration/repair in the inner ear. Curr. Opin. Neurobiol. 8: 480-487.

Staecker,H., Zheng,Q.Y., and Van De Water,T.R. 2001. Oxidative stress in aging in the C57B16/J mouse cochlea. Acta Otolaryngol. 121: 666-672.

Stallings,R.L., Ford,A.F., Nelson,D., Torney,D.C., Hildebrand,C.E., and Moyzis,R.K. 1991. Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics 10: 807-815.

Steel,K.P. 1998. Progress in progressive hearing loss. Science 279: 1870-1871.

Steel,K.P. 2000. Science, medicine, and the future: New interventions in hearing impairment. BMJ 320: 622-625.

Steel,K.P. and Kros,C.J. 2001. A genetic approach to understanding auditory function. Nat. Genet. 27: 143-149.

Stenberg,A.E., Wang,H., Sahlin,L., and Hultcrantz,M. 1999. Mapping of estrogen receptors alpha and beta in the inner ear of mouse and rat. Hear. Res. 136: 29-34.

Street,V.A., Robinson,L.C., Erford,S.K., and Tempel,B.L. 1995. Molecular genetic analysis of distal mouse chromosome 6 defines gene order and positions of the deafwaddler and opisthotonos mutations. Genomics 29: 123-130.

Street,V.A., McKee-Johnson,J.W., Fonseca,R.C., Tempel,B.L., and Noben-Trauth,K. 1998. Mutations in a plasma membrane Ca2+-ATPase gene cause deafness in deafwaddler mice. Nat. Genet. 19: 390-394.

Su,W., Jackson,S., Tjian,R., and Echols,H. 1991. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev. 5: 820-826.

Sud,R., Jones,C.M., Banfi,S., and Dawson,S.J. 2005. Transcriptional regulation by Barhl1 and Brn-3c in organ-of-Corti-derived cell lines. Brain Res. Mol Brain Res. 141: 174-180.

Sugano,K., Nakashima,Y., Yamaguchi,K., Fukayama,N., Maekawa,M., Ohkura,H., Kakizoe,T., and Sekiya,T. 1996. Sensitive detection of loss of heterozygosity in the TP53 gene in pancreatic adenocarcinoma by fluorescence-based single-strand conformation polymorphism analysis using blunt-end DNA fragments. Genes Chromosomes. Cancer 15: 157-164.

Suh,Y. and Vijg,J. 2005. SNP discovery in associating genetic variation with human disease phenotypes. Mutat. Res. 573: 41-53.

Suske,G. 1999. The Sp-family of transcription factors. Gene 238: 291-300.

Suske,G. 2000. Transient transfection of Schneider cells in the study of transcription factors. Methods Mol Biol. 130: 175-187.

Swinscow.T.D.V. and Campbell.M.J., 2002. Statistics at square one. London: BMJ Publishing Group.

Syvanen,A.C. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat. Rev. Genet. 2: 930-942.

Tabor,H.K., Risch,N.J., and Myers,R.M. 2002. Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat. Rev. Genet. 3: 391-397.

Tadokoro,K., Hashimoto,R., Tatsumi,M., Kamijima,K., and Kunugi,H. 2004. Analysis of enhancer activity of a dinucleotide repeat polymorphism in the neurotrophin-3 gene and its association with bipolar disorder. Neuropsychobiology 50: 206-210.

Takeuchi,S., Ando,M., and Kakigi,A. 2000. Mechanism generating endocochlear potential: role played by intermediate cells in stria vascularis. Biophys. J. 79: 2572-2582.

TAYLOR,W., PEARSON,J., MAIR,A., and BURNS,W. 1965. STUDY OF NOISE AND HEARING IN JUTE WEAVING. J. Acoust. Soc. Am. 38: 113-120.

Theil,T., McLean-Hunter,S., Zornig,M., and Moroy,T. 1993. Mouse Brn-3 family of POU transcription factors: a new aminoterminal domain is crucial for the oncogenic activity of Brn-3a. Nucleic Acids Res. 21: 5921-5929.

Theil,T., Zechner,U., Klett,C., Adolph,S., and Moroy,T. 1994. Chromosomal localization and sequences of the murine Brn-3 family of developmental control genes. Cytogenet. Cell Genet. 66: 267-271.

Thiel,G. and Cibelli,G. 2002. Regulation of life and death by the zinc finger transcription factor Egr-1. J. Cell Physiol 193: 287-292.

Thomas,G.R., Faulkes,D.J., Gascoyne,D., and Latchman,D.S. 2004. EWS differentially activates transcription of the Brn-3a long and short isoform mRNAs from distinct promoters. Biochem. Biophys. Res. Commun. 318: 1045-1051.

Thottassery,J.V., Sun,D., Zambetti,G.P., Troutman,A., Sukhatme,V.P., Schuetz,E.G., and Schuetz,J.D. 1999. Sp1 and egr-1 have opposing effects on the regulation of the rat Pgp2/mdr1b gene. J. Biol. Chem. 274: 3199-3206.

Tong,B., Grimes,H.L., Yang,T.Y., Bear,S.E., Qin,Z., Du,K., El Deiry,W.S., and Tsichlis,P.N. 1998. The Gfi-1B proto-oncoprotein represses p21WAF1 and inhibits myeloid cell differentiation. Mol Cell Biol. 18: 2462-2473.

Trieu,M., Ma,A., Eng,S.R., Fedtsova,N., and Turner,E.E. 2003. Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a. Development 130: 111-121.

Tsuprun,V., Goodyear,R.J., and Richardson,G.P. 2004. The structure of tip links and kinocilial links in avian sensory hair bundles. Biophys. J. 87: 4106-4112.

Unal,M., Tamer,L., Dogruer,Z.N., Yildirim,H., Vayisoglu,Y., and Camdeviren,H. 2005. N-acetyltransferase 2 gene polymorphism and presbycusis. Laryngoscope 115: 2238-2241.

Usami,S., Hjelle,O.P., and Ottersen,O.P. 1996. Differential cellular distribution of glutathione--an endogenous antioxidant--in the guinea pig inner ear. Brain Res. 743: 337-340.

Usami,S., Takumi,Y., Fujita,S., Shinkawa,H., and Hosokawa,M. 1997. Cell death in the inner ear associated with aging is apoptosis? Brain Res. 747: 147-150.

Vahava,O., Morell,R., Lynch,E.D., Weiss,S., Kagan,M.E., Ahituv,N., Morrow,J.E., Lee,M.K., Skvorak,A.B., Morton,C.C., Blumenfeld,A., Frydman,M., Friedman,T.B., King,M.C., and Avraham,K.B. 1998. Mutation in transcription factor POU4F3 associated with inherited progressive hearing loss in humans. Science 279: 1950-1954.

Van Camp,G., Willems,P.J., and Smith,R.J. 1997. Nonsyndromic hearing impairment: unparalleled heterogeneity. Am. J. Hum. Genet. 60: 758-764.

Van Campen,L.E., Murphy,W.J., Franks,J.R., Mathias,P.I., and Toraason,M.A. 2002. Oxidative DNA damage is associated with intense noise exposure in the rat. Hear. Res. 164: 29-38.

Van Eerdewegh,P., Little,R.D., Dupuis,J., Del Mastro,R.G., Falls,K., Simon,J., Torrey,D., Pandit,S., McKenny,J., Braunschweiger,K., Walsh,A., Liu,Z., Hayward,B., Folz,C., Manning,S.P., Bawa,A., Saracino,L., Thackston,M., Benchekroun,Y., Capparell,N., Wang,M., Adair,R., Feng,Y., Dubois,J., FitzGerald,M.G., Huang,H., Gibson,R., Allen,K.M., Pedan,A., Danzig,M.R., Umland,S.P., Egan,R.W., Cuss,F.M., Rorke,S., Clough,J.B., Holloway,J.W., Holgate,S.T., and Keith,T.P. 2002. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 418: 426-430.

Van Hauwe,P., Coucke,P.J., Ensink,R.J., Huygen,P., Cremers,C.W., and Van Camp,G. 2000. Mutations in the KCNQ4 K+ channel gene, responsible for autosomal dominant hearing loss, cluster in the channel pore region. Am. J. Med. Genet. 93: 184-187.

Van Laer,L., Huizing,E.H., Verstreken,M., van Zuijlen,D., Wauters,J.G., Bossuyt,P.J., Van de,H.P., McGuirt,W.T., Smith,R.J., Willems,P.J., Legan,P.K., Richardson,G.P., and Van Camp,G. 1998. Nonsyndromic hearing impairment is associated with a mutation in DFNA5. Nat. Genet. 20: 194-197.

Van Laer,L., DeStefano,A.L., Myers,R.H., Flothmann,K., Thys,S., Fransen,E., Gates,G.A., Van Camp,G., and Baldwin,C.T. 2002. Is DFNA5 a susceptibility gene for age-related hearing impairment? Eur. J. Hum. Genet. 10: 883-886.

Van Laer,L., Carlsson,P.I., Ottschytsch,N., Bondeson,M.L., Konings,A., Vandevelde,A., Dieltjens,N., Fransen,E., Snyders,D., Borg,E., Raes,A., and Van Camp,G. 2006. The contribution of genes involved in potassium-recycling in the inner ear to noise-induced hearing loss. Hum. Mutat. 27: 786-795.

van Leeuwen,H.C., Strating,M.J., Rensen,M., de Laat,W., and Van der Vliet,P.C. 1997. Linker length and composition influence the flexibility of Oct-1 DNA binding. EMBO J. 16: 2043-2053.

Verpy,E., Leibovici,M., Zwaenepoel,I., Liu,X.Z., Gal,A., Salem,N., Mansour,A., Blanchard,S., Kobayashi,I., Keats,B.J., Slim,R., and Petit,C. 2000. A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. Nat. Genet. 26: 51-55.

Verrijzer,C.P. and Van der Vliet,P.C. 1993. POU domain transcription factors. Biochim. Biophys. Acta 1173: 1-21.

Viguera,E., Canceill,D., and Ehrlich,S.D. 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. 20: 2587-2595.

Wallis,D., Hamblen,M., Zhou,Y., Venken,K.J., Schumacher,A., Grimes,H.L., Zoghbi,H.Y., Orkin,S.H., and Bellen,H.J. 2003. The zinc finger transcription factor Gfi1, implicated in lymphomagenesis, is required for inner ear hair cell differentiation and survival. Development 130: 221-232.

Wang,J., Pitarque,M., and Ingelman-Sundberg,M. 2006. 3'-UTR polymorphism in the human CYP2A6 gene affects mRNA stability and enzyme expression. Biochem. Biophys. Res. Commun. 340: 491-497.

Wang,S.W., Gan,L., Martin,S.E., and Klein,W.H. 2000. Abnormal polarization and axon outgrowth in retinal ganglion cells lacking the POU-domain transcription factor Brn-3b. Mol Cell Neurosci. 16: 141-156.

Wang,S.W., Mu,X., Bowers,W.J., Kim,D.S., Plas,D.J., Crair,M.C., Federoff,H.J., Gan,L., and Klein,W.H. 2002. Brn3b/Brn3c double knockout mice reveal an unsuspected role for Brn3c in retinal ganglion cell axon outgrowth. Development 129: 467-477.

Wang,W.Y., Barratt,B.J., Clayton,D.G., and Todd,J.A. 2005. Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 6: 109-118.

Wang,Y., Hirose,K., and Liberman,M.C. 2002. Dynamics of noise-induced cellular injury and repair in the mouse cochlea. J. Assoc. Res. Otolaryngol. 3: 248-268.

Warner,S.J., Hutson,M.R., Oh,S.H., Gerlach-Bank,L.M., Lomax,M.I., and Barald,K.F. 2003. Expression of ZIC genes in the development of the chick inner ear and nervous system. Dev. Dyn. 226: 702-712.

Weaver,D.T. and DePamphilis,M.L. 1982. Specific sequences in native DNA that arrest synthesis by DNA polymerase alpha. J. Biol. Chem. 257: 2075-2086.

Wegner,M., Drolet,D.W., and Rosenfeld,M.G. 1993. POU-domain proteins: structure and function of developmental regulators. Curr. Opin. Cell Biol. 5: 488-498.

Weinmaster,G. 1998. Notch signaling: direct or what? Curr. Opin. Genet. Dev. 8: 436-442.

Weiss,S., Gottfried,I., Mayrose,I., Khare,S.L., Xiang,M., Dawson,S.J., and Avraham,K.B. 2003. The DFNA15 deafness mutation affects POU4F3 protein stability, localization, and transcriptional activity. Mol Cell Biol. 23: 7957-7964.

Westin,G. and Schaffner,W. 1988. Heavy metal ions in transcription factors from HeLa cells: Sp1, but not octamer transcription factor requires zinc for DNA binding and for activator function. Nucleic Acids Res. 16: 5771-5781.

White,M.B., Carvalho,M., Derse,D., O'Brien,S.J., and Dean,M. 1992. Detecting single base substitutions as heteroduplex polymorphisms. Genomics 12: 301-306.

Wikman,F.P., Katballe,N., Christensen,M., Laurberg,S., and Orntoft,T.F. 2000. Efficient mutation detection in mismatch repair genes using a combination of single-strand conformational polymorphism and heteroduplex analysis at a controlled temperature. Genet. Test. 4: 15-21.

Wiley,T.L., Cruickshanks,K.J., Nondahl,D.M., and Tweed,T.S. 1999. Aging and middle ear resonance. J. Am. Acad. Audiol. 10: 173-179.

Willott,J.F., Turner,J.G., Carlson,S., Ding,D., Seegers,B.L., and Falls,W.A. 1998. The BALB/c mouse as an animal model for progressive sensorineural hearing loss. Hear. Res. 115: 162-174.

Wolfe,S.A., Nekludova,L., and Pabo,C.O. 2000. DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. 29: 183-212.

Wood,J.D., Muchinsky,S.J., Filoteo,A.G., Penniston,J.T., and Tempel,B.L. 2004. Low endolymph calcium concentrations in deafwaddler2J mice suggest that PMCA2 contributes to endolymph calcium maintenance. J. Assoc. Res. Otolaryngol. 5: 99-110.

Wood.K.J., 1998. The chemistry of bioluminescent reporter assays. Promega Notes 65: 14-20.

Wright,A., Charlesworth,B., Rudan,I., Carothers,A., and Campbell,H. 2003. A polygenic basis for late-onset disease. Trends Genet. 19: 97-106.

Wu,T., Ikezono,T., Angus,C.W., and Shelhamer,J.H. 1994. Characterization of the promoter for the human 85 kDa cytosolic phospholipase A2 gene. Nucleic Acids Res. 22: 5093-5098.

Xiang,M., Zhou,L., Peng,Y.W., Eddy,R.L., Shows,T.B., and Nathans,J. 1993. Brn-3b: a POU domain gene expressed in a subset of retinal ganglion cells. Neuron 11: 689-701.

Xiang,M., Zhou,L., Macke,J.P., Yoshioka,T., Hendry,S.H., Eddy,R.L., Shows,T.B., and Nathans,J. 1995. The Brn-3 family of POU-domain factors: primary structure, binding specificity, and expression in subsets of retinal ganglion cells and somatosensory neurons. J. Neurosci. 15: 4762-4785.

Xiang,M., Gan,L., Zhou,L., Klein,W.H., and Nathans,J. 1996. Targeted deletion of the mouse POU domain gene Brn-3a causes selective loss of neurons in the brainstem and trigeminal ganglion, uncoordinated limb movement, and impaired suckling. Proc. Natl. Acad. Sci. U. S. A 93: 11950-11955.

Xiang,M., Gan,L., Li,D., Chen,Z.Y., Zhou,L., O'Malley,B.W., Jr., Klein,W., and Nathans,J. 1997. Essential role of POU-domain factor Brn-3c in auditory and vestibular hair cell development. Proc. Natl. Acad. Sci. U. S. A 94: 9445-9450.

Xiang,M., Gao,W.Q., Hasson,T., and Shin,J.J. 1998. Requirement for Brn-3c in maturation and survival, but not in fate determination of inner ear hair cells. Development 125: 3935-3946.

Xiao,W. and Oefner,P.J. 2001. Denaturing high-performance liquid chromatography: A review. Hum. Mutat. 17: 439-474.

Yamada,N., Yamaya,M., Okinaga,S., Nakayama,K., Sekizawa,K., Shibahara,S., and Sasaki,H. 2000. Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema. Am. J. Hum. Genet. 66: 187-195.

Yamashita,D., Jiang,H.Y., Schacht,J., and Miller,J.M. 2004. Delayed production of free radicals following noise exposure. Brain Res. 1019: 201-209.

Yamoah,E.N., Lumpkin,E.A., Dumont,R.A., Smith,P.J., Hudspeth,A.J., and Gillespie,P.G. 1998. Plasma membrane Ca2+-ATPase extrudes Ca2+ from hair cell stereocilia. J. Neurosci. 18: 610-624.

Yang,M., Tan,H., Zheng,J.R., Wang,F., Jiang,C., He,M., Chen,Y., and Wu,T. 2006. [Association of cadherin CDH23 gene polymorphisms with noise induced hearing loss in Chinese workers]. Wei Sheng Yan. Jiu. 35: 19-22.

Yang,T.T. and Chow,C.W. 2003. Transcription cooperation by NFAT.C/EBP composite enhancer complex. J. Biol. Chem. 278: 15874-15885.

Yang,Y., Hwang,C.K., Junn,E., Lee,G., and Mouradian,M.M. 2000. ZIC2 and Sp3 repress Sp1-induced activation of the human D1A dopamine receptor gene. J. Biol. Chem. 275: 38863-38869.

Ye,S. 2000. Polymorphism in matrix metalloproteinase gene promoters: implication in regulation of gene expression and susceptibility of various diseases. Matrix Biol. 19: 623-629.

Yip,S.P., Hopkinson,D.A., and Whitehouse,D.B. 1999. Improvement of SSCP analysis by use of denaturants. Biotechniques 27: 20-2, 24.

Yu,H., Song,Q., Freedman,B.I., Chao,J., Chao,L., Rich,S.S., and Bowden,D.W. 2002. Association of the tissue kallikrein gene promoter with ESRD and hypertension. Kidney Int. 61: 1030-1039.

Zelko,I.N., Mariani,T.J., and Folz,R.J. 2002. Superoxide dismutase multigene family: a comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. Free Radic. Biol. Med. 33: 337-349.

Zhang,B., Ye,S., Herrmann,S.M., Eriksson,P., de Maat,M., Evans,A., Arveiler,D., Luc,G., Cambien,F., Hamsten,A., Watkins,H., and Henney,A.M. 1999. Functional polymorphism in the regulatory region of gelatinase B gene in relation to severity of coronary atherosclerosis. Circulation 99: 1788-1794.

Zhao,C. and Meng,A. 2005. Sp1-like transcription factors are regulators of embryonic development in vertebrates. Dev. Growth Differ. 47: 201-211.

Zheng,J., Shen,W., He,D.Z., Long,K.B., Madison,L.D., and Dallos,P. 2000a. Prestin is the motor protein of cochlear outer hair cells. Nature 405: 149-155.

Zheng,J.L. and Gao,W.Q. 2000. Overexpression of Math1 induces robust production of extra hair cells in postnatal rat inner ears. Nat. Neurosci. 3: 580-586.

Zheng,J.L., Shou,J., Guillemot,F., Kageyama,R., and Gao,W.Q. 2000b. Hes1 is a negative regulator of inner ear hair cell differentiation. Development 127: 4551-4560.

Zheng,Q.Y. and Johnson,K.R. 2001. Hearing loss associated with the modifier of deaf waddler (mdfw) locus corresponds with age-related hearing loss in 12 inbred strains of mice. Hear. Res. 154: 45-53.

Zheng,Q.Y., Yan,D., Ouyang,X.M., Du,L.L., Yu,H., Chang,B., Johnson,K.R., and Liu,X.Z. 2005. Digenic inheritance of deafness caused by mutations in genes encoding cadherin 23 and protocadherin 15 in mice and humans. Hum. Mol Genet. 14: 103-111.

Zheng,Y., Ikeda,K., Nakamura,M., and Takasaka,T. 1998. Endonuclease cleavage of DNA in the aged cochlea of Mongolian gerbil. Hear. Res. 126: 11-18.

Zine,A., Van De Water,T.R., and de Ribaupierre,F. 2000. Notch signaling regulates the pattern of auditory hair cell differentiation in mammals. Development 127: 3373-3383.

Zine,A., Aubert,A., Qiu,J., Therianos,S., Guillemot,F., Kageyama,R., and de Ribaupierre,F. 2001. Hes1 and Hes5 activities are required for the normal development of the hair cells in the mammalian inner ear. J. Neurosci. 21: 4712-4720.

Zine,A. and de Ribaupierre,F. 2002. Notch/Notch ligands and Math1 expression patterns in the organ of Corti of wild-type and Hes1 and Hes5 mutant mice. Hear. Res. 170: 22-31.

Zoghbi,H.Y. and Orr,H.T. 2000. Glutamine repeats and neurodegeneration. Annu. Rev. Neurosci. 23: 217-247.

Zweidler-McKay,P.A., Grimes,H.L., Flubacher,M.M., and Tsichlis,P.N. 1996. Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. Mol Cell Biol. 16: 4024-4034.

APPENDIX A

Royal Free and University College Medical School
UNIVERSITY COLLEGE LONDON

Centre for Auditory Research, UCL Ear Institute.

INFORMATION SHEET.....VERSION 4

Title: Investigating adult onset hearing loss.

You are being invited to take part in a research study to investigate the causes of hearing loss in adults. Participation in the study will **not** involve you making additional visits to the hospital, it will take about 30 minutes on the day of your appointment. Before you decide whether you would like to take part it is important for you to read this sheet which explains what the research is for and what it involves. Please take your time and ask if you need more information or explanation. If you wish to then you can also take time to discuss your decision with others. Thank you for reading this.

Consumers for Ethics in Research (CERES) publish a leaflet entitled "Medical Research and You". This leaflet gives more information about medical research and looks at some questions you may want to ask. A copy may be obtained from CERES,

- **What is the purpose of the study ?**

The Molecular Audiology Research Group at University College London is looking for volunteers to take part in an on-going study to investigate the causes of adult onset hearing loss. Hearing loss is very common as people get older; a third of people over sixty have some degree of hearing loss. We are looking to see whether people who lose their hearing do so because they have inherited an increased susceptibility to hearing loss. This work will help researchers to identify if some people are at greater risk of developing adult onset hearing loss, and pinpoint the reasons as to why this happens. This line of work would pave the way for future preventative therapies to prevent or halt the onset of adult hearing loss.

- **Why have I been chosen?**

You have suffered some hearing loss as an adult so you fit the criteria to take part in the study. It is hoped that we will eventually have several hundred patients taking part in the study.

- **Do I have to take part?**

No. This project is entirely voluntary, if you do not wish to take part then please say so. You do not have to give any reason for deciding not to take part and your decision will not affect the standard of care your receive. If you do decide to take part you will be asked to sign a consent form. You will then be given a copy of the consent form and this information sheet to keep. You can chose to withdraw at any time without giving a reason.

- **What do we need from you if you do take part?**

If you choose to volunteer for this project we will take a small blood sample from you (5mls, about a teaspoonful). This will be taken as part of your outpatient's visit at the hearing aid clinic; no extra visits to the hospital are necessary to take part in the study. You won't need to provide anything else or undergo any further tests for this project. The only other thing we need is for you to fill in a short questionnaire giving us some basic information about you & your hearing loss. One question will be regarding your ethnic background. It is important that we obtain this information for our study. This is because susceptibility to adult onset hearing loss could vary amongst different ethnic groups. We estimate that all this will take less than 30 minutes. A copy of your hearing test will also be supplied to the research study.

- **Will this information be confidential?**

Yes. This information will be strictly confidential; it will only be used as part of this research project and will not be given to any third parties. The research project is carried out with each sample being given a reference number rather than a patient name. Your personal details including your name and address will not be disclosed outside this hospital. Therefore you will not be contacted with specific results of the project. You will not be identified in any scientific report or publication relating to this study. However, if you are interested in the general outcome of the project this will be available.

What will happen to my blood sample?

Genetic testing will be performed on your blood sample with the other samples from people who are losing their hearing. At the same time we will be performing genetic testing on samples from people in the general population who have not lost their hearing. If we find any differences between the two groups it will help us to identify the cause of adult onset hearing loss. Your sample may also be used in other hearing related studies following approval by a research ethics committee.

- **Are there any advantages/disadvantages to taking part in the study?**

Apart from helping research into hearing loss you will not receive any benefit from taking part. The disadvantage of taking part is having a sample of blood drawn and spending some time filling in our form.

- **Who is funding this research and who will be performing the work?**

The project is being funded by a charity called Research into Ageing who are part of Help the Aged. The Molecular Audiology Group at University College London headed by Dr Sally Dawson will perform the research study. Lisa Nolan will undertake the main research work as part of her PhD. Lisa herself is hearing impaired and has a keen interest in hearing research.

- **Where can I get more information?**

If you wish to know more information about the project in general or wish to find out the results of the research you can contact the research group leader: Dr Sally Dawson on or by email at

Or, Lisa Nolan on (minicom) (via BT Type-talk) or by email:

This project has been reviewed and approved by The Royal Free Hospital and Medical School Local Research Ethics Committee.

Thank you for taking the time to read this information sheet, and thank you if you decide to take part in this study. Our research would not be possible without volunteers.

Royal Free and University College Medical School
UNIVERSITY COLLEGE LONDON

INSTITUTE OF LARYNGOLOGY AND OTOLOGY
Audiology Division

CAR

Dr. Barbara Cadge, FRCP
Audiological Physician

QUESTIONNAIRE II: Adult Onset Hearing Loss	Patient number:	

Thank you for helping with this project which will contribute to the understanding of the nature of hearing loss in advancing years. Please see the attached Information Sheet for more details. We are looking for people of any age who have developed a hearing loss as an adult. Your contribution will involve answering the few questions below, and having a single blood sample taken which will be sent for genetic analysis.

1. At what age were you first aware of hearing problems?

<20 years ☐ 20-40 years ☐ 41-60 years ☐
61-80 years ☐ >81 years ☐

2. Does anyone else in your family have hearing problems? Yes ☐ No ☐

If yes, please give details.----------

---

3. Do you think your hearing loss was due to exposure to excessive noise?

Yes ☐ No ☐

If yes, please give details.

Royal National Throat, Nose and Ear Hospital

THE QUEEN'S AWARD FOR TECHNOLOGICAL ACHIEVEMENT 1998

Website: www.ucl.ac.uk/audiological-science

4. Have you had any other significant medical or ear problems? Yes ☐ No ☐

If yes, please give details,

---

5. How old are you? ☐ years

6. Please indicate which of the following ethnic group best describes you. If you are of mixed ethnicity please tick all those groups which are relevant to you.

White	Asian or Asian British	Black or Black British
☐ British or Irish	☐ Indian	☐ Caribbean
☐ Jewish	☐ Pakistan	☐ African
☐ North Europe	☐ Bangladeshi	☐ Somali
☐ South Europe	☐ East African Asian	☐ Other, please specify
☐ Other, please specify	☐ Other, please specify	

Other

☐ Chinese ☐ Japanese ☐ Arab/Iranian/ North African

Other Please, specify

Thank you for your time and attention. Your co-operation has been much appreciated.

Royal Free and University College Medical School
UNIVERSITY COLLEGE LONDON

INSTITUTE OF LARYNGOLOGY AND OTOLOGY
Audiology Division

Dr. Barbara Cadge, FRCP
Audiological Physician

UCL Centre for Auditory Research

Patient Number: ☐

CONSENT FORM

Title of Project: **Adult Onset Hearing Loss**

Name of Researcher: Dr Sally Dawson

Please initial box

1. I confirm that I have read and understand the information sheet dated .....version..... for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐
3. I understand that sections of any of my medical notes may be looked at by responsible individuals or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. ☐
4. I agree to take part in the above study. ☐

Name of Patient	Date	Signature
Name of Person taking consent (if different from researcher)	Date	Signature
Researcher	Date	Signature

Royal National Throat, Nose and Ear Hospital

1998
THE QUEEN'S AWARD FOR TECHNOLOGICAL ACHIEVEMENT

Website: www.ucl.ac.uk/audiological-science

APPENDIX B

Reference sequence for exon 1 of Brn-3c (human): NCBI GenBank Accession U10060.1.
Reference sequence for exon 2 of Brn-3c (human): NCBI GenBank Accession U10061.1.
Reference sequence for intron of Brn-3c (human): NCBI GenBank Accession AF043452.1.

Reference sequence for human Brn-3c promoter: Ensembl transcript ID ENST00000230732). Promoter sequence is shown on next page and genetic variation identified and confirmed as part of the PCR-SSCP screen is highlighted: -3495$(GT)_{19}$: black bold (NCBI SNP Cluster ID rs28994882), -3457$(GA)_{1}$: red bold (NCBI SNP Cluster ID rs28994881), -3432$(G)_{11}$: black bold (NCBI SNP Cluster ID rs29001168), -1391A: black bold (NCBI SNP Cluster ID rs1368402), -566$(GT)_{21}$: black bold (NCBI SNP Cluster ID rs28987086), -386C: black bold (NCBI SNP Cluster ID rs28994880). The numbering above the sequence is with respect to the first A of the ATG translation start as +1.

Reference sequence for human Brn-3c promoter: Ensembl transcript IDENST00000230732).

```
....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3740      -3730      -3720      -3710      -3700      -3690
ctgga ggatatggtg ctttgtgccc tgacttcacc ttacatgtgg caaatctatg tccct

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3680      -3670      -3660      -3650      -3640      -3630
actag actctgagct cctgaagcag tgtcttctgt atcttcaagg ccttccacaa gactc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3620      -3610      -3600      -3590      -3580      -3570
tgctg ttagtaggta tttggaaatg cttatcgaat ggggaaaatt cttccttcaa tgcca

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3560      -3550      -3540      -3530      -3520      -3510
cttgc ataaatgccc acatagtcct tacatggacc ttcacatatt ttgtttggcc cataa

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3500      -3490      -3480      -3470      -3460      -3450
atggt atga[gtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtga]ttgtaa tttaa

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3440      -3430      -3420      -3410      -3400      -3390
tgcca tggtggt[ggg ggggggggg]tg ggtaagtgct ctctagttta ccacagtttc aatct

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3380      -3370      -3360      -3350      -3340      -3330
ttcct tttatccacc cagataaagt tgctcaattt tattttacta gttttgcacc ttata

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3320      -3310      -3300      -3290      -3280      -3270
gaatt taagtttaca aattctggaa agcatagagg ctgttacaaa ttgccgtcta aggct

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3260      -3250      -3240      -3230      -3220      -3210
gtcct tccaattgct ttttctatgg tcttttccct cttttcttct ccagtctcca gtctg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3200      -3190      -3180      -3170      -3160      -3150
ttatc tccaaggcct tctcaaaaag actggtcctt aaggatgggt gttgtcttac cgctg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3140      -3130      -3120      -3110      -3100      -3090
gctgc ctagaactgc cagtagacct tcttagaggg gaagatggga aaaaccctaa aacac

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3080      -3070      -3060      -3050      -3040      -3030
atggg accagacaca ggtgtgtttc ttgatgtgtg atcctaatgc ctggggatgc tggag
```

```
....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -3020      -3010      -3000      -2990      -2980      -2970
taggg gctggtaagg tgaaggtcaa cattggaggc atgtgaggac cagttagttt tgctc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2960      -2950      -2940      -2930      -2920      -2910
tgttt ctgactcatg cctgacctct atgctggcca gtatctccca gacccctgag aggac

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2900      -2890      -2880      -2870      -2860      -2850
agagc agaaaaagag aatggatatg aatcctttct ctcccagttt ctaggggttc agctt

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2840      -2830      -2820      -2810      -2800      -2790
cacag acaactaggg tattatggaa tgattatggg tctcagggca tcatgggacc aggca

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2780      -2770      -2760      -2750      -2740      -2730
aaata cagaccagtg tatccctgtg tttcattccc cttttccctt tgggcttcca gcatt

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2720      -2710      -2700      -2690      -2680      -2670
gggaa gaaaacaaag actggtctta aatcctagtg agtgataaaa tactgcattc agctc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2660      -2650      -2640      -2630      -2620      -2610
tgggg agttgagtgt ctctggagat aaattcctga catgctcaaa gtatgtttca ttcaa

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2600      -2590      -2580      -2570      -2560      -2550
attaa ccagaatgat agtcttgtta aagctggaag tcttgttaaa gcttagcaat catct

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2540      -2530      -2520      -2510      -2500      -2490
tatac agcccactcc ttgtaggtaa actgaggtcc agtcagggtc acccatactc cgtgt

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2480      -2470      -2460      -2450      -2440      -2430
cagaa tgcagcccca aatccagccc tcctgcagcc catagttctg atgtgtattc tgcac

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2420      -2410      -2400      -2390      -2380      -2370
ttcta gaatcctgcc acccttatgc tctgtgtctt gttcatgaat ctacctccct tctac

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2360      -2350      -2340      -2330      -2320      -2310
tgctt ggcaagttta gagttaactt tctgaagtga atattagact ggtgactcct tgagt

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2300      -2290      -2280      -2270      -2260      -2250
gccag gatggattct gattcacatc tgtgtcccag cctttggcat ggggtctggc acaat
```

```
....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2240      -2230      -2220      -2210      -2200      -2190
gcaga tgttgggtga atggagagca aggatccaca gtatgcacgt gcagacaaac cgggt

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2180      -2170      -2160      -2150      -2140      -2130
tccag gaggatgact tggcccaccc ccactctcac atgctttctc tttaagttga ggaag

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2120      -2110      -2100      -2090      -2080      -2070
aagag aggaacatgt attcaatgat atcaatgcgc caggcactgt gccatgtgct tatcc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2060      -2050      -2040      -2030      -2020      -2010
cagtg cattcctcac aacaatcaat gggaaataaa tgatcacccc cattttacag atggg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -2000      -1990      -1980      -1970      -1960      -1950
gacac cgcagtacag agaagttgag aatatctcgt ggtctcagcc agcaagaggc agagg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1940      -1930      -1920      -1910      -1900      -1890
cgcat tttgaatctg aggccagggc ctgagcgcta tggcaccagg ttgtgaggga gttta

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1880      -1870      -1860      -1850      -1840      -1830
ctgag caccaatttc gtggctcaat gcttttcttc ttaccctaac tctgaacgtc ttcac

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1820      -1810      -1800      -1790      -1780      -1770
cagaa tccaaggagg cagacagtag tgcaattcca tttaagagcc gccgaaaccg aggct

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1760      -1750      -1740      -1730      -1720      -1710
cagag ggctttagcg aaagggttgc gcttggctct aggcacctgg gctgagtctt tttcc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1700      -1690      -1680      -1670      -1660      -1650
cacca tcgcagggcg ggaaccagcg gaggggggctg gctcggatgg ggagaaaagc aactg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1640      -1630      -1620      -1610      -1600      -1590
gaggg cgccgagggg aagagggagc ccggatctgt cagggcgtcc tcttggacta aggga

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1580      -1570      -1560      -1550      -1540      -1530
tgttc ccctaaacca caccacccca cctcgttcag attctgggaa acccggcacg cacat

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1520      -1510      -1500      -1490      -1480      -1470
accct gcacaataac aggcagggtg aggtctctaa gccccgaaga tcgctctctc agcgg
```

```
....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1460      -1450      -1440      -1430      -1420      -1410
aggca gtggccctga cgaggggtga cagctgcacg cgcggggttt ctctccgaac cggag

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1400      -1390      -1380      -1370      -1360      -1350
tgcag cgtagtcgag gtccaggatt ccccatccat tattcacgat gtttactaga gcggg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1340      -1330      -1320      -1310      -1300      -1290
gcaag ggagaggaaa gaagagggaa gtagagagaa aaagaggcta ggggaggtgg aggcg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1280      -1270      -1260      -1250      -1240      -1230
ggcag tgctaacctc gagagccctc aagttccgaa actttgagaa ggaagaccaa gaggc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1220      -1210      -1200      -1190      -1180      -1170
taagg cgcctgggaa gcagcaggcc gtcagtaaat atttgtaaga tggatggata actgg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1160      -1150      -1140      -1130      -1120      -1110
gtgag tgagtgaatg ggctaatcat tagcccctct gatccttgtt ttcctcatct gtaaa

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1100      -1090      -1080      -1070      -1060      -1050
acggg ataacatcac ctactccata cggtgggtat gagaatccca tgggcaaact ctaaa

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
  -1040      -1030      -1020      -1010       -1000       -990
ttgtc tagcacatat taggaactct taaacggtag ctgttgtcac gaggaatggg cttgc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -980       -970       -960       -950       -940       -930
aagct ctggctgccc agcagaaatg cactgaagga ctcccaggcc tggagggcca tcctg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -920       -910       -900       -890       -880       -870
aacag tcgctattct aaaaaaaata ctccacaagc ttccttagac gggagggttg gagga

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -860       -850       -840       -830       -820       -810
agagt gggccaagtt acatccctca aatgaaagaa agaaaggcaa gcgtggggag gaagc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -800       -790       -780       -770       -760       -750
caccc cggactgaga aggcagttac tgcccctacc tctaccccgg agcgcggttg aggga

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -740       -730       -720       -710       -700       -690
ggtgg ggcaggggtc acctgggcct cgttctggca gcccctcacc ctctccaggg cccgt
```

```
....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -680       -670       -660       -650       -640       -630
ctggg cgcttggagg cgcctcctcg ctgcgccgcg ggaccggact ctggtggaca gcttg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -620       -610       -600       -590       -580       -570
ggcgt gaggccagga gcgccctgga aatgggcagt ttggcggcag ccgggccgac ggagt

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -560       -550       -540       -530       -520       -510
gtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtggtgg agaggggaag ttgga

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -500       -490       -480       -470       -460       -450
gtagg gtcaacttcc tgccccagct cagcccaggg ctacccttt atccaggcag ttcga

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -440       -430       -420       -410       -400       -390
gctgg gacaggacgg agaggttggg acttttgggg tggcatgggg gaagggaagt ccacg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -380       -370       -360       -350       -340       -330
aagaa gaaagaatcg gaaaggtctg gcgggttgga gccagcgggg cggggcggac tggga

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -320       -310       -300       -290       -280       -270
gaggc caggccaggc ccgggtataa aggctgtgga gggggcggcc gccgcgggcg cagaa

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -260       -250       -240       -230       -220       -210
aggcg cgccgctagc tgctgtctct cctcacctcc cgggccgccc ctgcgagtcc ccggc

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -200       -190       -180       -170       -160       -150
gcgtg agcacgcctg cgcgcgcccg ggcccttcct ggcaggctgc ttgtaagatg agtga

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -140       -130       -120       -110       -100       -90
agaag caggtggggg agaggggagg cagcgagcga gagggcgagg ggagcgcggg cgctg

....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
   -80        -70        -60        -50        -40        -30
agcag cgctcacttg gagagcggca agcaagctag acaagcctga ttccatgtca cccgc

....| ....|....| ....|....
   -20        -10
tgcca ccctgccagg agcgcgaag
```

APPENDIX C.

Table A: Results of bioinformatic analysis using MatInspector software (www.genomatix.de/products/MatInspector/index.html) to identify putative transcription factor binding sites within the region spanning the -3432poly-G polymorphism. Results are shown for each -3432poly-G allele sequence as 'core' and 'matrix' similarity scores using the following settings: core similarity set to 0.75 and matrix similarity set to optimised; Matrix family library version 3.1.1. For poly-G allele sequence see Table 3.2 section 3.2.18.

		SNPG1		SNPG2		SNPG4		SNPG5		$(G)_{10}$		$(G)_{11}$		$(G)_{12}$	
Family	**Matrix (version)**	Core	Matrix	Core	Matrix	Core	Matrix	Core	Matrix	Core	Matrix	Core	Matrix	Core	Matrix
ZPBF	**ZBP-89 (01)**	1.000	0.946	1.000	0.950					1.000	0.973	1.000	0.984	1.000	0.984
ZBPF	**ZF9 / CPBP (01)**			0.923	0.884							0.820	0.872	0.820	0.876
EGRF	**Egr-4 / NGFI-C (01)**			0.787	0.843										
EGRF	**Egr-1 / NGFI-A (01)**	1.000*	0.942*	0.797	0.808										
EGRF	**WT1 (01)**	0.796	0.893	0.796	0.864					0.796	0.884	0.796	0.884	0.796	0.884
INSM	**INSM1 (01)**	1.000	0.909	1.000	0.950	1.000	0.909			1.000	0.909	1.000	0.909	1.000	0.909
SP1F	**SP1 (01)**	0.796	0.899	0.796	0.893			1.000	0.958	0.796	0.906	0.796	0.910	0.796	0.919
MAZF	**MAZR (01)**	1.000	0.943	1.000	0.952	1.000	0.925	1.000	0.885	1.000	0.957	1.000	0.957	1.000	0.957
MAZF	**MAZ (01)**							0.866	0.900						
SPIF	**GC box elements (01)**	0.819	0.909	0.819	0.891	0.819	0.882	1.000	0.929	0.819	0.908	0.819	0.907	0.819	0.904
RREB	**RREB1 (01)**	1.000	0.831	1.000	0.853	1.000	0.833			1.000	0.853	1.000	0.853	1.000	0.853
SP1F	**TIEG (01)**	1.000	0.868	1.000	0.858	1.000	0.858	1.000	0.858	1.000	0.858	1.000	0.858	1.000	0.858
GLIF	**ZIC2 (01)***			1.000	0.948	1.000	0.948			1.000	0.948	1.000	0.948	1.000	0.948

* Results were obtained during preparation of this thesis using Matrix family library version 5.0.

APPENDIX D

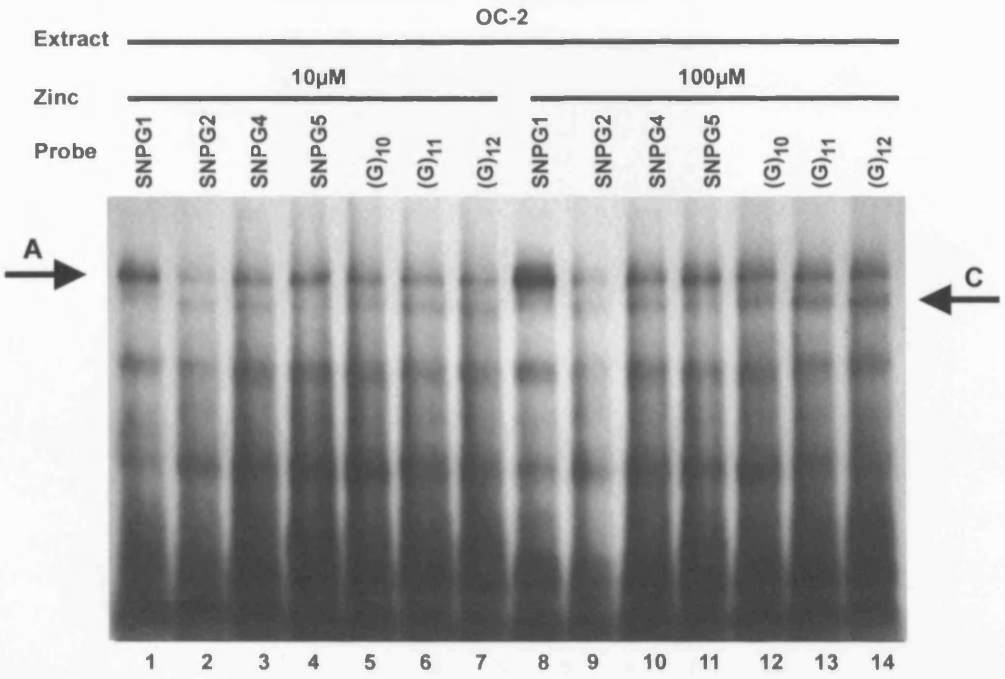


Figure 5.10 Effect of zinc on binding of nuclear proteins A and C to common -3432 poly-G alleles. Probes: SNPG1 (lanes 1 & 8), SNPG2 (lanes 2 & 9), SNPG4 (lanes 3 & 10), SNPG5 (lanes 4 & 11) $(G)_{10}$ (lanes 5 & 12), $(G)_{11}$ (lanes 6 & 13) or $(G)_{12}$ (lanes 7 &14) were incubated with 8μg OC-2 nuclear extract in the presence of 10μM zinc (lanes 1-7) or 100μM zinc (lanes 8-14). Arrows A and C denote the location of shifted nuclear protein-DNA complexes A and C, respectively.

APPENDIX E

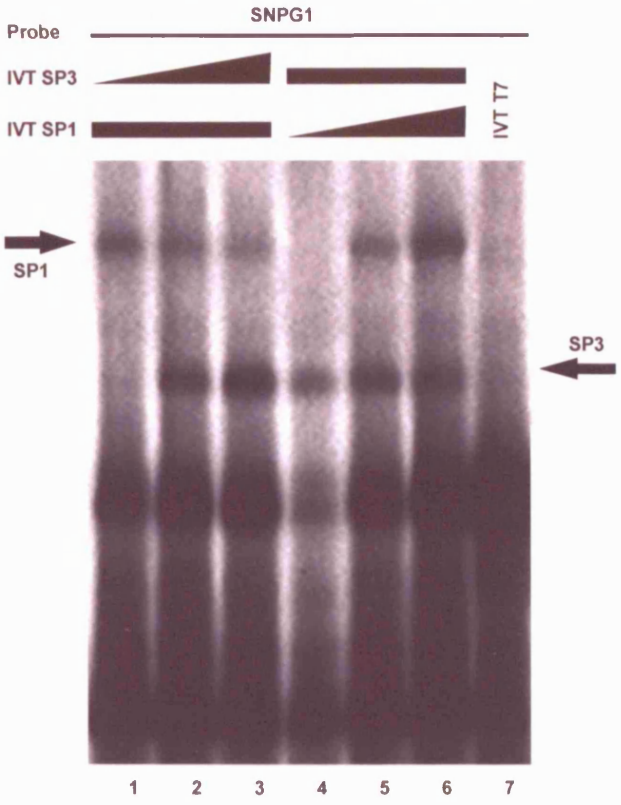


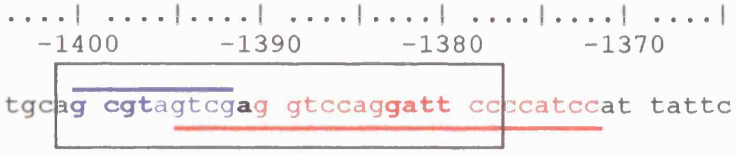
Figure 5.15c Effect of -3432poly-G polymorphism on binding of SP3 and SP1. The SNPG1 probe was incubated in the presence of 3.5µl SP1 *in-vitro* translate (lanes 1-3) in the presence of increasing amounts of SP3 *in-vitro* translate: 1µl (lane 2) and 2µl (lane 3); 1µl SP3 *in-vitro* translate (lanes 4-6) in the presence of increasing amounts of SP1 *in-vitro* translate: 3.5µl (lane 5) and 7µl (lane 6) or 8µl T7 *in-vitro* translate as a negative control for endogenous protein binding.

APPENDIX F.

(i) Table B: Results of bioinformatic analysis using MatInspector software (www.genomatix.de/products/MatInspector/index.html) to identify putative transcription factor binding sites within the region spanning the -1391A>C single nucleotide substitution. Results are shown for each allele sequence as 'core' and 'matrix' similarity scores using the following settings: core similarity set to 0.75 and matrix similarity set to optimised; Matrix family library version 3.1.1. For -1391 allele sequence see Table 3.2 section 3.2.18.

		-1391A		-1391C	
Family	**Matrix (version)**	Core	Matrix	Core	Matrix
EGRF	**EGR3 (01)**	1.000	0.797	1.000	0.796
GFI1	**GFI1 (01)**	1.000	0.873	1.000	0.873

(ii) Section of Brn-3c promoter spanning -1391A>C. The single nucleotide substitution is highlighted in black bold. Boxed area size of oligonucleotide used in EMSA analysis. Sequence used by MatInspector software to predict putative transcription factor binding sites is underlined and highlighted: red denotes sequence used to predict the putative Gfi-1 binding site (the core sequence is in red bold), blue denotes sequence used to predict the putative Egr3 binding site (the core sequence is in blue bold).



APPENDIX G

Sequencing of cloned PCR products as a confirmatory method for genotyping the -3432poly-G polymorphism by direct sequencing.

Sequencing of cloned PCR products to elucidate and / or to confirm discrete poly-G alleles identified by direct sequencing was not a useful approach. 'Stutter' alleles from the PCR amplification confounded discrimination of the genuine alleles and hence, confirmation of genotype.

'Stutter' alleles are contractions and / or expansions of a repetitive DNA sequence and are thought to arise due to DNA polymerase slippage when repetitive DNA sequences are amplified by PCR (Clarke et al, 2001; Shinde et al, 2003). Repetitive DNA sequences including both mononucleotide and dinucleotide repeats have an intrinsic propensity to DNA polymerase slippage; during normal replication slipped-strand mispairing results in short deletions and / or additions in repeat length (Levinson and Gutman, 1987a; Gragg et al, 2002; for review, see Kunkel and Bebenek, 2000). It is thought that the ability of repetitive DNA sequences to readily adopt secondary conformations results in a tendency for progression of the ongoing DNA polymerase to stall (Weaver and DePamphilis, 1982; Bichara et al, 2000; Viguera et al, 2001 and references within). 'Slippage' of the DNA polymerase during a PCR reaction is thought to involve a transient dissociation of the template and the 3' end of the primer strand followed by mis-pairing within the repeat unit upon annealing termed, 'slipped strand mis-pairing' which results in an extrahelical loop composed of the unpaired repeat units (Viguera et al, 2001 and references within). If this loop forms on the 'template strand' and persists until the next round of DNA amplification, contraction of the repeat unit occurs. Conversely, if the loop forms on the 'primer strand' then expansion of the repeat unit occurs (Levisnon and Gutman, 1987a).

Both contraction and expansion of the -3432poly-G repeat length was observed during analysis of cloned PCR products. For example, presence of the stutter allele $(G)_9$ (contraction) was observed for subjects identified $(G)_{11}$ / $(G)_{10}$ and the presence of the stutter allele $GGC(G)_{11}$ (expansion) was observed for subjects identified SNPG2 / SNPG6 [$GGC(G)_9$ / $GGC(G)_{10}$]. In general, the usefulness of sequencing cloned PCR products to aid elucidation and / or to confirm discrete poly-G alleles was very much dependent on the nature of the poly-G alleles present proving more useful for homozygous samples than heterozygous. For subjects homozygous for the -3432poly-G polymorphism it was possible to confirm genuine alleles by sequencing of cloned PCR products; this was particularly evident in the case of subjects homozygous for allele SNPG1, [$GGC(G)_8$], the shortest allele length identified that contained a single nucleotide substitution within the poly-G repeat. Sequencing of individual clones (n = 10) from the PCR product pool did not reveal the presence of any stutter alleles for subjects SNPG1 homozygous. In contrast, for subjects heterozygous for the -3432poly-G polymorphism sequencing of cloned PCR products was not very informative, especially if the longest allele combinations were present. For example, for subjects identified $(G)_{12}$ / $(G)_{11 \text{ or } 10}$, sequencing of cloned PCR products was confounded by stutter alleles of differing poly-G repeat length. The abundance of stutter alleles observed for some samples that contain long poly-G allele combinations and lack of stutter alleles for subjects that contain short poly-G alleles is consistent with the concept that the frequency of DNA polymerase slippage is reported to increase as the number of repeating units in a repetitive DNA sequence increases (Sia et al, 1997; Clarke et al, 2001). In addition, the degree of stutter alleles identified and the origin of the multi-

allelic -3432poly-G polymorphism may in part be explained by the tendency of mononucleotide runs of poly-G/C to be very susceptible to DNA polymerase slippage (Gragg et al, 2002); probably to some extent due to strong stabilisation of secondary conformations in sequences of poly-G/C that facilitate slippage (Sagher et al, 1999).

The AmpliTaq Gold (Promega) DNA polymerase used to amplify the -3432poly-G polymorphism by PCR prior to either direct sequencing or sequencing of individual clones of the PCR product pool does not have proofreading capability. It is possible that use of a DNA polymerase with proofreading capacity may have generated more informative results from sequencing of cloned PCR products by reducing the occurrence of stutter alleles in the final product pool. However, it has been reported by Clarke et al, 2001 that *Taq* and the proofreading polymerase *Pfu* (high fidelity) generate similar errors at mononucleotide and dinucleotide repeats. It should also be considered that some of the stutter alleles observed upon sequencing of cloned PCR products may not have originated during the PCR amplification but rather during propagation of the pGEM®-T easy plasmids in the *E.Coli* strain used for sub-cloning, XL1-Blue (Stratagene). Indeed, clones containing microsatellite sequence inserts can vary in size due to mutations that arise during propagation of the plasmid in E. Coli (Levinson and Gutman, 1987b; Freund et al, 1989; Bichara et al, 2000). However, the *E.coli* strain used, XL-1, carry the *recA* mutation, which is known to dramatically reduce the frequency of homologous recombination and hence, improve insert stability.

In summary, sequencing cloned PCR products as a secondary method to direct sequencing to elucidate and / or to confirm discrete poly-G alleles was not a useful approach; the presence of stutter alleles made discrimination of genuine poly-G alleles extremely difficult.

APPENDIX H

Elucidation of problematic -3432poly-G genotypes

SNPG4 homozygotes.

Originally a total of ten subjects were assigned to chromatogram group 8 based on direct sequencing analysis. From observation of the sequencing chromatogram typical for this group of subjects (Fig. H.1a) it appears clear that the genotype is homozygous; no convincing heterozygous peaks are observed across the poly-G repeat or downstream invariant region. Consequently, a genotype of SNPG4 homozygous was determined.

To confirm this genotype alleles were sized by PAGE analysis of radioactive PCR products for each of the ten subjects. The results of this secondary analysis revealed that only three subjects exhibited a single band of 182bp equivalent to a genotype of SNPG4 homozygous (for example see Fig. H.1c, lane 1). The remaining seven subjects in addition to exhibiting the 182bp band also displayed an extra band of 183bp, which due to having a strong intensity could not be discarded as a stutter allele from the PCR reaction (for example see Fig. H.1c, lanes 2 and 3). Repeat PAGE analysis of radioactive PCR products confirmed these observations. Based on these results the three subjects that displayed a single band of 182bp were verified as SNPG4 homozygous and the seven subjects that displayed the extra 183bp band were assigned a different genotype; SNPG4 hereozygous where the second allele was **n**ot **d**etermined (SNPG4 / **ND**).

It was not possible to clarify the genotype of subjects identified SNPG4 / ND accordingly this category of subjects was excluded from subsequent frequency analysis It is possible that the 183bp band is a PCR artefact; DNA polymerase slippage or the tendency of *Taq* DNA polymerase to randomly append the 3' end of DS DNA fragments with an extra residue, typically adenine, are two mechanisms that could explain the secondary band differing in length by an extra residue (Hauge et al, 1993). It is also feasible that this secondary band does not differ in length to that of the 182bp band but rather adopts an alternative secondary conformation such that migration is retarded by a factor of 1bp (Litt et al, 1993). Alternatively, it is possible that these subjects are heterozygous for the SNPG4 allele in which case allele $(G)_{12}$ having a length equivalent to 183bp is a likely candidate for the second allele. The only other poly-G allele equivalent to 183bp is SNPG2 but, SNPG2 / SNPG4 heterozygotes exhibit very clear and distinct results upon direct sequencing (see Fig. H.1b). In addition, it cannot be ruled out that a novel allele may exist in this group of subjects. In either case, it is possible that the SNPG4 allele shows preferential amplification over the second allele type during the PCR reaction hence, leading to the production of a direct sequencing chromatogram typical of SNPG4 homozygotes.

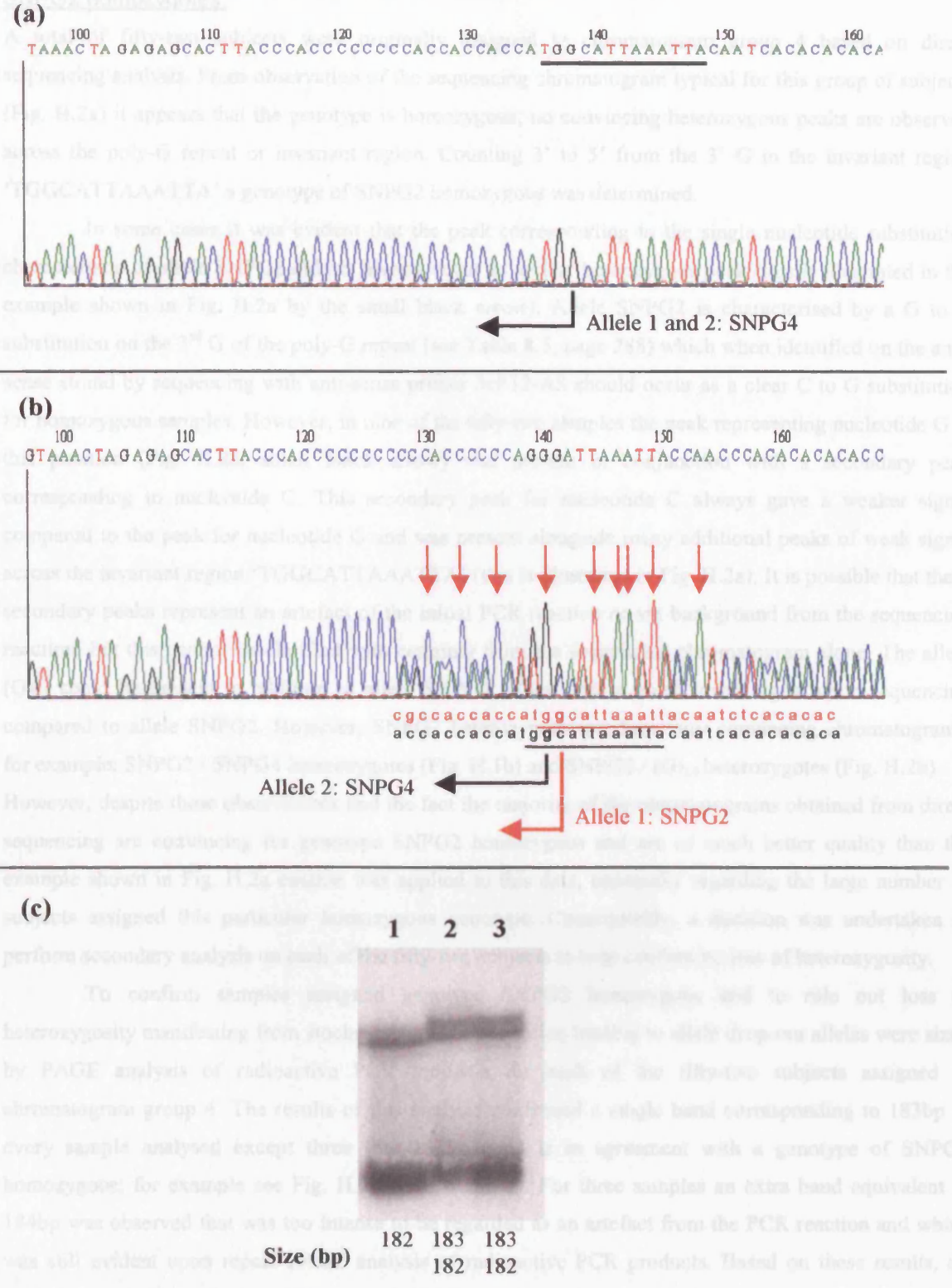


Figure H.1 Determination of SNPG4 homozygotes. **(a)** and **(b)** Direct sequencing chromatograms; data obtained is for the anti-sense strand. Alleles were elucidated by reading 3' to 5' from the 3' G within the invariant region (underlined). **(a)** Direct sequencing chromatogram representative of subjects assigned to chromatogram group 8 (genotype: SNPG4 homozygous). **(b)** Direct sequencing chromatogram representative of subjects assigned to chromatogram group 15 (genotype: SNPG2 / SNPG4 heterozygous). Clear distinction between homozygous and heterozygous peaks facilitated identification of each allele; see written sequence added below the chromatogram (homozygous peaks are indicated by the small red arrows). **(c)** Sizing of -3432poly-G alleles by PAGE analysis of radioactive PCR products using primers G3cP13-S and 3cP12-AS in the PCR. Lane 1: genotype SNPG4 homozygous, lanes 2 and 3: genotype SNPG4 / ND (second allele: Not Determined).

SNPG2 homozygotes.

A total of fifty-two subjects were originally assigned to chromatogram group 4 based on direct sequencing analysis. From observation of the sequencing chromatogram typical for this group of subjects (Fig. H.2a) it appears that the genotype is homozygous; no convincing heterozygous peaks are observed across the poly-G repeat or invariant region. Counting 3' to 5' from the 3' G in the invariant region 'TGGCATTAAATTA' a genotype of SNPG2 homozygous was determined.

In some cases it was evident that the peak corresponding to the single nucleotide substitution characteristic of allele SNPG2 did not always occur as a clear homozygous peak (this is illustrated in the example shown in Fig. H.2a by the small black arrow). Allele SNPG2 is characterised by a G to C substitution on the 3rd G of the poly-G repeat (see Table 8.5, page 288) which when identified on the anti-sense strand by sequencing with anti-sense primer 3cP12-AS should occur as a clear C to G substitution for homozygous samples. However, in nine of the fifty-two samples the peak representing nucleotide G at this position (Fig. H.2a, small black arrow) was present in conjunction with a secondary peak corresponding to nucleotide C. This secondary peak for nucleotide C always gave a weaker signal compared to the peak for nucleotide G and was present alongside many additional peaks of weak signal across the invariant region 'TGGCATTAAATTA' (this is illustrated in Fig. H.2a). It is possible that these secondary peaks represent an artefact of the initial PCR reaction or are background from the sequencing reaction, but this cannot be clarified with certainty from the sequencing chromatogram alone. The allele $(G)_{11}$ could be present in addition to allele SNPG2, albeit with a much lower signal upon sequencing compared to allele SNPG2. However, SNPG2 heterozygotes produce very convincing chromatograms, for example: SNPG2 / SNPG4 heterozygotes (Fig. H.1b) and SNPG2 / $(G)_{10}$ heterozygotes (Fig. H.2b). However, despite these observations and the fact the majority of the chromatograms obtained from direct sequencing are convincing for genotype SNPG2 homozygous and are of much better quality than the example shown in Fig. H.2a caution was applied to this data, especially regarding the large number of subjects assigned this particular homozygous genotype. Consequently, a decision was undertaken to perform secondary analysis on each of the fifty-two subjects to help confirm no loss of heterozygosity.

To confirm samples assigned genotype SNPG2 homozygous and to rule out loss of heterozygosity manifesting from stochastic hypoamplification leading to allele drop-out alleles were sized by PAGE analysis of radioactive PCR products for each of the fifty-two subjects assigned to chromatogram group 4. The results of this analysis confirmed a single band corresponding to 183bp in every sample analysed except three (the 183bp band is in agreement with a genotype of SNPG2 homozygous; for example see Fig. H.2c lanes 4 and 7). For three samples an extra band equivalent to 184bp was observed that was too intense to be regarded as an artefact from the PCR reaction and which was still evident upon repeat PAGE analysis of radioactive PCR products. Based on these results, all subjects assigned to chromatogram group 4 were confirmed as SNPG2 homozygous except the three displaying the extra band. The three subjects not confirming to the SNPG2 homozygous genotype were excluded from subsequent frequency analysis.

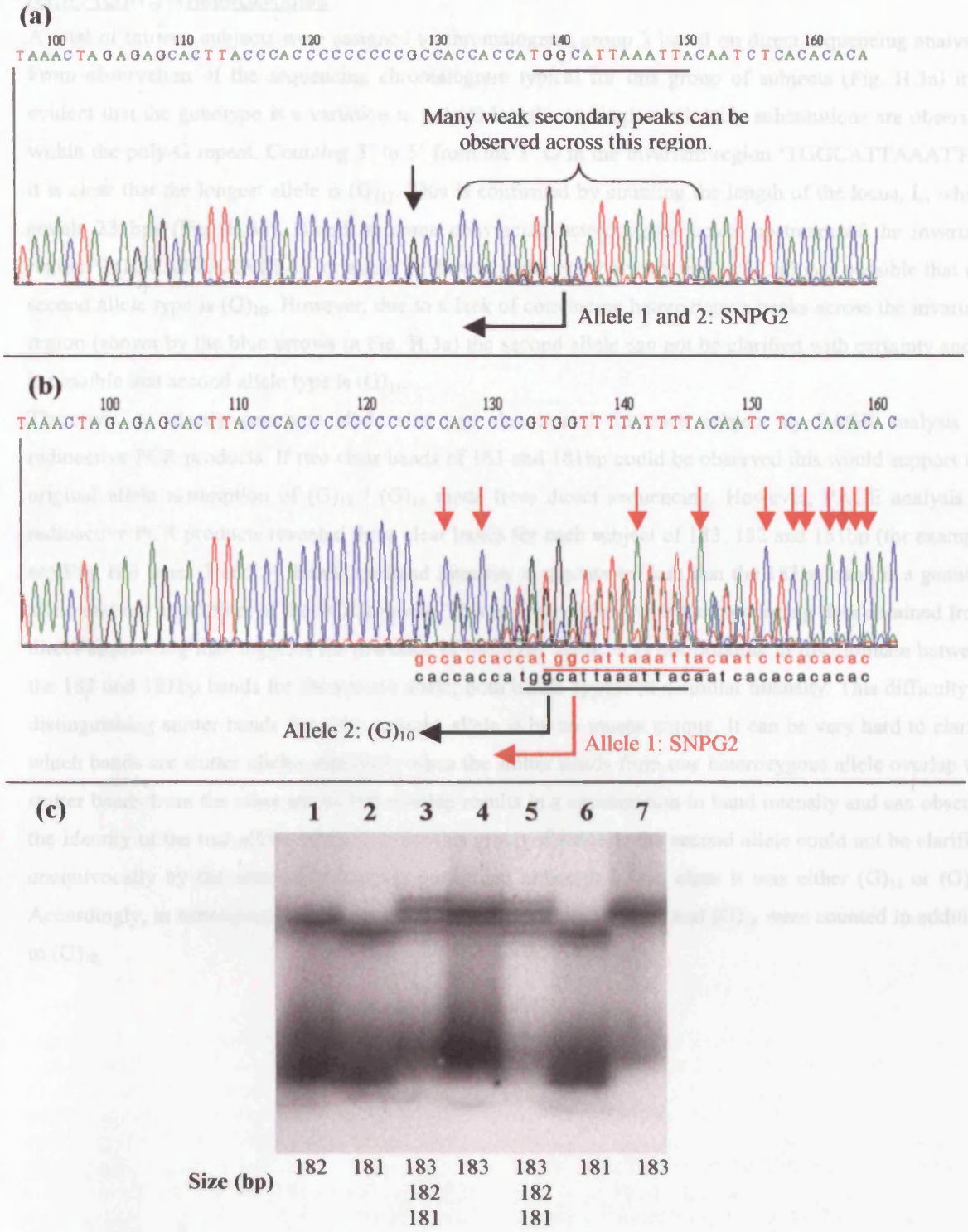
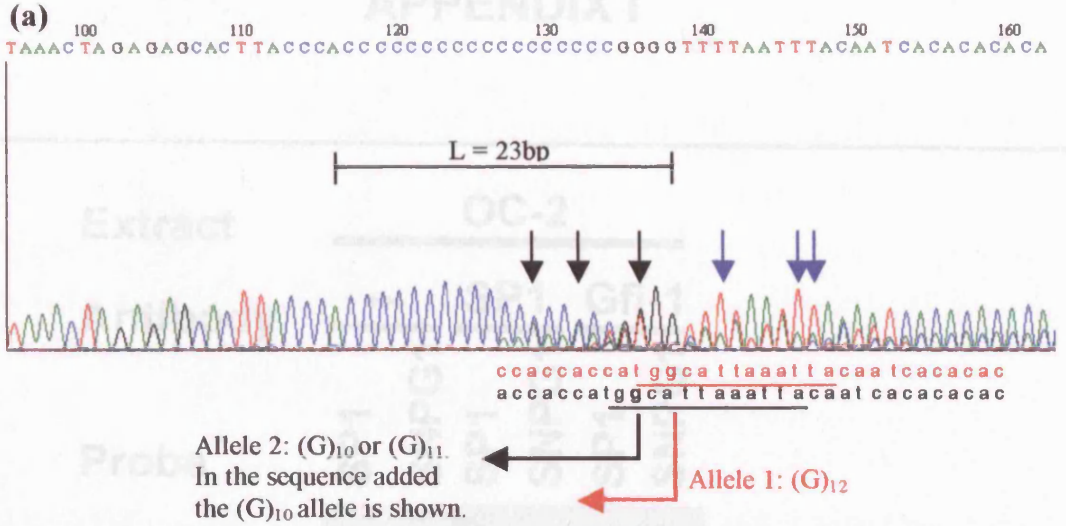


Figure H.2 Determination of SNPG2 homozygotes. **(a)** and **(b)** Direct sequencing chromatograms; data obtained is for the anti-sense strand. Alleles were elucidated by reading 3' to 5' from the 3' G within the invariant region (underlined). **(a)** Direct sequencing chromatogram representative of subjects assigned to chromatogram group 4 (genotype: SNPG2 homozygous). Example shown shows presence of many weak secondary peaks (small black arrow and region as indicated) that confounded assignment of SNPG2 homozygous. **(b)** Direct sequencing chromatogram representative of subjects assigned to chromatogram group 5 (genotype: SNPG2 / $(G)_{10}$ heterozygous). Clear distinction between homozygous and heterozygous peaks facilitated identification of each allele; see written sequence added below the chromatogram (homozygous peaks are indicated by the small red arrows). **(c)** Sizing of -3432poly-G alleles by PAGE analysis of radioactive PCR products using primers G3cP13-S and 3cP12-AS in the PCR. Lanes 1 & 2: 182 and 181bp size markers, respectively. Size markers were generated from pGEM®-T easy constructs carrying either alleles $(G)_{11}$ or $(G)_{10}$ as template in the PCR reaction. Lanes 3 & 5: genotype $(G)_{12}$ / $(G)_{11 \text{ or } 10}$, lanes 4 & 7 genotype: SNPG2 homozygous and lane 6: genotype $(G)_{10}$ homozygous.

$(G)_{12}$ / $(G)_{11 \text{ or } 10}$ heterozygotes

A total of thirteen subjects were assigned to chromatogram group 3 based on direct sequencing analysis. From observation of the sequencing chromatogram typical for this group of subjects (Fig. H.3a) it is evident that the genotype is a variation in poly-G length; no single nucleotide substitutions are observed within the poly-G repeat. Counting 3' to 5' from the 3' **G** in the invariant region 'TGGCATTAAATTA' it is clear that the longest allele is $(G)_{12}$. This is confirmed by counting the length of the locus, L, which equals 23 bps (Fig. H.3a). Based on some convincing heterozygous peaks upstream of the invariant region 'TGGCATTAAATTA' for allele 1 (shown by the black arrows in Fig. H.3a) it is possible that the second allele type is $(G)_{10}$. However, due to a lack of convincing heterozygous peaks across the invariant region (shown by the blue arrows in Fig. H.3a) the second allele can not be clarified with certainty and it is possible that second allele type is $(G)_{11}$.

Therefore, to clarify genotype allele size was determined for each subject by PAGE analysis of radioactive PCR products. If two clear bands of 183 and 181bp could be observed this would support the original allele assumption of $(G)_{12}$ / $(G)_{10}$ made from direct sequencing. However, PAGE analysis of radioactive PCR products revealed three clear bands for each subject of 183, 182 and 181bp (for example see Fig. H.3 lanes 3 and 5). Based on band intensity it appears evident that the 183bp band is a genuine allele and not an artefact of the PCR process. This is in agreement with the convincing data obtained from direct sequencing that suggests the presence of the $(G)_{12}$ allele. It is not possible to discriminate between the 182 and 181bp bands for the second allele; both bands appear of a similar intensity. This difficulty in distinguishing stutter bands from the genuine allele is by no means unique. It can be very hard to clarify which bands are stutter alleles especially when the stutter bands from one heterozygous allele overlap the stutter bands from the other allele; this overlap results in a combination in band intensity and can obscure the identity of the true allele. Therefore, for this group of subjects the second allele could not be clarified unequivocally by the secondary analysis performed although it was clear it was either $(G)_{11}$ or $(G)_{10}$. Accordingly, in subsequent frequency analysis both possibilities $(G)_{11}$ and $(G)_{10}$ were counted in addition to $(G)_{12}$.



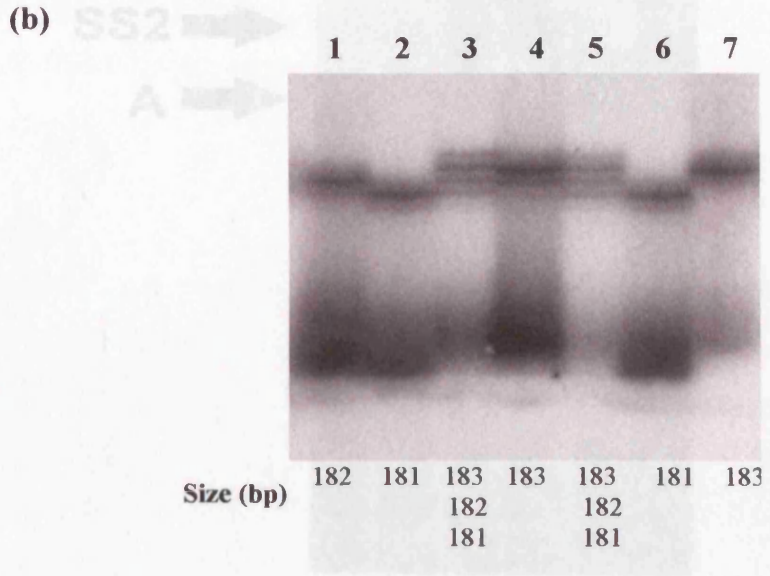
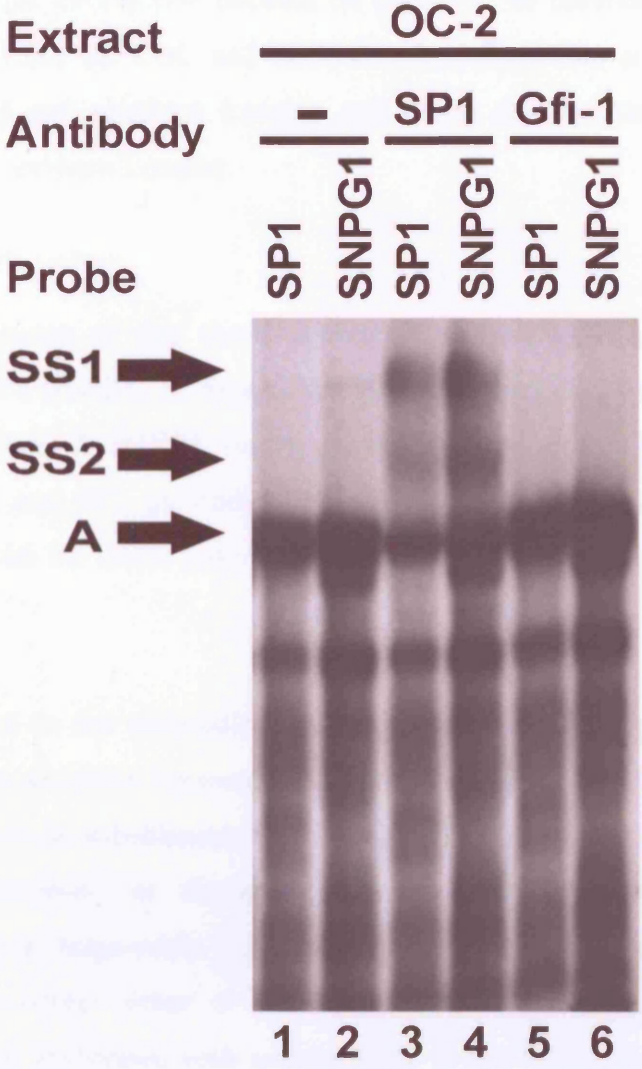


Figure H.3 Determination of $(G)_{12}$ / $(G)_{11 \text{ or } 10}$. **(a)** Direct sequencing chromatogram representative of subjects assigned to chromatogram group 3 (genotype: $(G)_{12}$ / $(G)_{11 \text{ or } 10}$); data shown is for the anti-sense strand. Alleles were elucidated by reading 3' to 5' from the 3' G within the invariant region (underlined). Allele 1 is $(G)_{12}$ as confirmed by the length of the locus L. Based on clear heterozygous peaks (small black arrows) it is possible the second allele is $(G)_{10}$ but due to the presence of unconvincing heterozygous peaks as indicated by the small blue arrows this cannot be confirmed; $(G)_{11}$ is also a possibility. **(b)** Sizing of -3432poly-G alleles by PAGE analysis of radioactive PCR products using primers G3cP13-S and 3cP12-AS in the PCR; this figure is reproduced from Fig. H.2c. Lanes 1 & 2: 182 and 181bp size markers, respectively. Size markers were generated from pGEM®-T easy constructs carrying either alleles $(G)_{11}$ or $(G)_{10}$ as template in the PCR reaction. Lanes 3 & 5: genotype $(G)_{12}$ / $(G)_{11 \text{ or } 10}$, lanes 4 & 7 genotype: SNPG2 homozygous and lane 6: genotype $(G)_{10}$ homozygous.

APPENDIX I



Supershift analysis confirms nuclear protein A is SP1. The labelled SP1 consensus sequence (lanes 1, 3 & 5) or SNPG1 probe (lanes 2, 4 & 6) were incubated with 8-10μg of OC-2 nuclear extract in the absence of antibody (lanes 1-2) or in the presence of anti-SP1 antibody (lanes 3-4) or anti-Gfi-1-1 antibody (lanes 5-6) as indicated. The location of nuclear protein-DNA complex A is denoted by the arrow, A. In the presence of the anti-SP1 antibody two shifted complexes SS1 & SS2 can be observed confirming SP1 is a component of nuclear protein-DNA complex A. Antibody was added subsequent to addition of probe and reactions were allowed to incubate for a further 30 minutes on ice. Antibodies were purchased from Santa Cruz Biotechnology, Inc.: rabbit polyclonal anti-SP1 antibody [SP1 (H-225) X] and goat polyclonal anti-Gfi-1 antibody [Gfi-1 (N-20) X].

Scientific Corrections.

Chapter 3.

The control sample group has been described previously (see Crawley et al, 1999). In addition to the information detailed on page 101 all control samples were Anglo Saxon Caucasians from the U.K. and were recruited from two sources: laboratory personnel from central and northern London and blood donors attending two general medical practices in northern London.

Chapter 5.

After submission of this thesis supershift analysis with a rabbit polyclonal anti-SP1 antibody [SP1 (H-225) X; Santa Cruz Biotechnology, Inc.] confirmed that OC-2 derived nuclear protein A is SP1 (work performed during the PhD project used a mouse monoclonal anti-SP1 antibody; SP1 (1C6) X, Santa Cruz Biotechnology, Inc.). For supershift with the rabbit polyclonal anti-SP1 antibody see Appendix I.

Chapter 8.

With respect to the association study it is acknowledged that the null hypothesis that there is no association between the -3432poly-G polymorphism and / or the -1391 A>C single nucleotide substitution and susceptibility to late onset hearing loss may actually be true. However, as discussed in section 8.8 (page 307) without for example, undertaking a large-scale population based association study it cannot be reliably concluded whether either or both of these common sequence variants in the Brn-3c promoter are associated with susceptibility to late onset hearing loss. This is especially so in light of the retrospective power calculation for the -1391 A>C single nucleotide substitution. Using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/#tdt_ins) it can be estimated that the power of the preliminary association study to detect a small genetic effect (relative risk: 1.4) is 24% using unscreened controls for the control sample group. Indeed, even to detect a larger genetic effect (relative risk: 2.0) the power of the preliminary association study is limited at approximately 68%. Sample sizes needed to detect a small genetic effect whilst maintaining at least 80% power is discussed further in section 8.8, page 309.